ESD-TR-87-061

2

Technical Report
781

Mathematical Foundations of Signal Processing
II. The Role of Group Theory

R.B. Holmes

13 October 1987

Lincoln Laboratory
MASSACHUSETTS INSTITUTE OF TECHNOLOGY
LEXINGTON, MASSACHUSETTS

Prepared for the Department of the Army
under Electronic Systems Division Contract F19628-85-C-0002.

Approved for public release; distribution unlimited.

The work reported in this document was performed at Lincoln Laboratory, a center for research operated by Massachusetts Institute of Technology, with the support of the Department of the Army under Air Force Contract F19628-85-C-0002.

This report may be reproduced to satisfy needs of U.S. Government agencies.

The views and conclusions contained in this document are those of the contractor and should not be interpreted as necessarily representing the official policies, either expressed or implied, of the United States Government.

The ESD Public Affairs Office has reviewed this report, and it is releasable to the National Technical Information Service, where it will be available to the general public, including foreign nationals.

This technical report has been reviewed and is approved for publication.

FOR THE COMMANDER

Arthur H. Wendel

Arthur H. Wendel, Captain, USAF
Acting Chief, ESD Lincoln Laboratory Project Office

Non-Lincoln Recipients

PLEASE DO NOT RETURN

Permission is given to destroy this document when it is no longer needed.

MASSACHUSETTS INSTITUTE OF TECHNOLOGY
LINCOLN LABORATORY

MATHEMATICAL FOUNDATION OF SIGNAL PROCESSING
II. THE ROLE OF GROUP THEORY

R.B. HOLMES
Group 32

TECHNICAL REPORT 781

13 OCTOBER 1987

LEXINGTON MASSACHUSETTS

ABSTRACT

Several aspects of group theory that prove useful for various signal processing applications are presented.

Chapter I begins with a discussion of signal processing activities and goals at an abstract level, and continues with a look at the mathematical underpinnings of this subject. There follows a list of specific mathematical results that seem to be of greatest relevance to signal processing.

Chapter II surveys the role played by infinite groups in modeling signals and filters. Here substantial use is made of the associated harmonic analysis, in the abelian case the dual group serves as the natural frequency domain.

Chapter III presents a fairly detailed review of the representation theory of finite groups, through the Plancherel formula. The essential idea here is to then use those special unitary transforms which are also group transforms for digital signal compression and decorrelation, and the associated group filters as fast suboptimal Wiener (or other) filters. Initial evidence suggests that nonabelian group filters can improve on the standard DFT/FFT methods without significant increase in computational complexity.

Chapters I and III are written at an elementary level for wide access; Chapter II is written at a higher level, requiring some background in functional and harmonic analysis. Comments are inserted throughout to suggest various generalizations of the material under discussion.

Chapter IV contains a summary of the main points and conclusions, and suggests some directions for further research, particularly on the use of finite nonabelian group transforms and filters.

TABLE OF CONTENTS

LIST OF ILLUSTRATIONS

LIST OF TABLES

MATHEMATICAL FOUNDATIONS OF SIGNAL PROCESSING II. THE ROLE OF GROUP THEORY

I. BACKGROUND TO SIGNAL PROCESSING

In this introductory chapter we will set down some general principles and philosophy of signal processing, while attempting to avoid the details of specific applications. The field is by now far too vast and multifaceted to permit any simple summary or encapsulation. Our aims will be modest: to agree on some terminology, some historical background, and some of the goals of signal processing. There follows a brief resume of the problems and methods of signal processing. Together this material is intended to furnish a compressed understanding of the field at an abstract level.

An inevitable consequence of an author's professional experience and personal predilections is a particular and usually subjective answer to the basic questions: What is interesting? What is important? In the present context this results in a neutral, mathematical approach, free of implementation considerations and the requirements of specific technologies. We are searching for mathematical paradigms of some elegance and widespread applicability. There is some analogy here with physical theories and formulas, where a single principle can be given a crisp mathematical formulation and then applied in a variety of real situations. There are also certain more specific analogies between physics and signal theory, based on a common underlying mathematical formulation. A familiar example is the uncertainty principle which, in its mathematical essence, is an instance of the local/global duality between Fourier transform pairs (see below, Section II.4). But such analogies should not be pressed too far, because the basic tools of signal processing (Fourier transform, bandlimited function, stationary process, etc.) are after all, mathematical and not empirical in nature. We have a greater freedom in selecting signal models than physical models; utility and efficiency are the major criteria, rather than agreement with experimental data.

This report represents a continuation of the author's interest in the foundations of signal processing. An earlier paper [1] provided a detailed operator-theoretic treatment of discrete-time single-channel signal processing. In the terminology of Section II.2 below such a signal would be described as a weakly stationary random process on the group of integers. Extensions of some of the basic results of [1] to modeling and filtering processes on more general groups are given in Sections II.2-II.4. Later, in 1983, the author gave a course "Fundamentals of Signal Processing" at the Lincoln Laboratory, which remains available on videotape. There the view was taken that a common goal of signal processing activity is to identify a model which in some sense explains a given pattern of observations. If the model is considered as an unknown element of a suitable Hilbert space, and is assumed related to the observations by a linear (and often compact) transformation, then estimates of the model can be constructed by operator-theoretic methods again; for example, by the use of singular vector expansions, pseudoinversion, and regularization.

The resulting estimates often have interpretation as various types of spline functions, and also as Bayes estimates for an appropriate prior. This methodology can also be viewed as yielding particular cases of general optimal algorithms [2].

In contrast to these purely Hilbert space techniques we are going to consider below the role played in signal processing by another fundamental mathematical structure — groups. In Chapter II we discuss several signal models that can be defined on and/or analyzed in terms of an underlying group structure. This material is presented as a rapid survey at a fairly advanced level. In Chapter III we take an opposite tack, and go into some more elementary material in greater detail. There the emphasis is on 'finite' — the application of finite groups to finite dimensional signal processing. We discuss the use of group transforms, especially those that are 'fast', for coding and pattern recognition purposes, and group filters, for signal estimation. This material is of real practical value and there are ample opportunities for further research. Our intent is primarily to expose the basic concepts and issues.

I.1 WHAT IS SIGNAL PROCESSING?

Let us begin with a provisional definition: a *signal* is the output of an array of sensors configured in time and/or space. Accordingly, then, a signal represents observations made of some physical process and, we presume, it carries information pertaining to the state of some physical system.

Most signals of engineering interest occur in the context of remote sensing. We will understand this term to refer, in a generic way, to observations made at a distance by devices sensitive to some sort of energy. This energy could be electromagnetic in nature (gamma and x-rays, visible light, infrared, radio and television, etc.), acoustical, or vibrational (mechanical, seismic) Remote sensing systems may be classified as *active* or *passive*, according as the received energy is that produced by a man made transmitter and then scattered by an object of interest, or is produced (or reflected) by the object alone.

Passive systems naturally occur in the contexts of astronomy, photography, satellite scanning, and geophysical recording. Of course, in a different, nonengineering, direction, we might include economic systems. Active systems include radar/sonar, a variety of medical imaging devices (CAT, NMR, PET, US, etc.), industrial procedures employing CAT-like equipment for quality testing purposes, often called nondestructive evaluation (NDE), and seismic prospecting.

Naturally, each of the above areas is a major field in itself, and so there is by now a massive literature in signal processing and its applications: many conference proceedings, including the annual ICASSP Proceedings from the IEEE, several specialized journals and 200 or so books. The present report is intended to have a small overlap with this literature.

It is possible to partition the history of signal processing in the 20th century into 3 eras, as indicated in the following table (I-1). Of course, such brevity cannot do justice to all the developments and authors involved in the signal processing business; our table is intended only to

TABLE I-1

Eras of Signal Processing in the 20th Century

1910- 1940	Physical:	Vacuum tubes, lumped circuits
	Analytical:	Impulse response, transfer function, transform methods
	Names:	Fourier/Laplace, Bode, Nyquist
1940- 1960	Physical:	Microwave circuits
	Analytical:	Statistical concepts (correlation, matched filters, information theory)
	Names:	Gabor, Shannon, Wiener
1960- Present	Physical:	Digitial computer (permits realization of arbitrary transfer functions), integrated circuits, optical technology
	Analytical:	Digital filters, spectrum estimation, fast Fourier transform, linear inverse theory
	Names:	Kailath, Oppenheim, Slepian, Tukey

be suggestive rather than complete. What we do want to stress is that current signal processing activity draws on many disciplines from within mathematics, as well as on computer science and integrated circuit technology. There is in general an ongoing dynamic interplay between algorithms and architecture. A broad survey of the field at present is given in the collection [4], edited by T. Kailath; again, there is little overlap with the present work.

Let us now restate our provisional definition of a signal as follows: the sensor output referred to at the beginning of this section will now be called *data* (often, measurements, observations, —). The phrase 'signal processing' will henceforth be replaced by *data processing*, and will be taken to mean the purposeful modification of data in order to eliminate redundancy or to extract information. Let us in turn take this last phrase to mean the construction (or, identification) of a mathematical model which, in some sense, 'explains' the data. Finally, we make the following definition: a *signal* is an unobservable mathematical quantity related to the data whose value can be inferred from an identified model.

What do we want to learn from data processing? A couple of very abstract goals were stated above. Somewhat more specifically we list the following possible goals as indicative of the major problems of the field:

- Data compression, decorrelation and feature extraction (efficient representation for display, storage, transmission, pattern recognition, etc.)
- Recognize significant aspects of the data (trends, periodicities, etc.)
- Signal detection
- Signal estimation
- Simulation (use of an identified model to generate more, artificial, data which, in some sense, behaves like the original data)

Now, in terms of a specific goal, one of the most interesting ensuing questions (subjectively speaking, again) concerns the nature of the actual operations that we elect to perform on the data. How are these to be justified; why do we do one thing and not another? Some of the factors which impinge on this decision are the following:

- goal of the processing (as above)
- model structure (physical or synthetic)
- prior information and constraints
- performance criterion
- nature and amount of data
- computational time available (circuit speed, necessity for real-time operation, etc.)

In a way, these factors collectively define an abstract paradigm for data processing, in that we may expect their specification to (eventually) result in the choice of a particular numerical algorithm.

We might offer two further comments about the design and use of algorithms. We have just indicated that many factors must be specified before an algorithm can possibly emerge. This is exactly the reason there is such a diversity of data processing algorithms extant, and why so often two of them are not directly comparable. Thus, when trying to select from among the legion of available algorithms, one's first task is to be sure that it was designed to achieve the user's goal and that it is consistent with the other factors just listed. There are other general desiderata also, when selecting an existing algorithm or attempting to design one's own. There should be a 'good question' in the background, that is, a well-defined or, more technically, a well-posed problem whose solution exists uniquely, for given data, and depends continuously on the data. The (approximate) solution should be efficiently computable and stable with respect to measurement errors ('noise'). (Even well-posed problems often result in numerically ill-conditioned equations as, for example, in the common case of computation of the pseudoinverse of a linear operator of finite rank. Naturally, ill conditioning will amplify any measurement noise present.)

These desiderata, along with that of rapid convergence as the associated discretization decreases, while giving guidance in algorithm design, also show that considerable care and effort are required. Thus, the first comment about algorithm design is that it is both hard to do well and hard to compare competing results.

The second comment pertains at a more philosophical level to this entire business of what might be termed the algorithmic-centered approach to engineering problems. In this approach, as just outlined, one designs an algorithm based on the physics and geometry of a particular observational situation, and in accord with the preceding general guidelines, then another algorithm is proposed and analyzed, and so forth. The existence of so many algorithms in the engineering literature suggests that this approach is indeed widespread.

An alternative, gradually acquiring some well deserved acceptance, is what might be termed the information-centered approach. Here one basically indicates the type and quantity of information (that is, models plus data) available about an unknown signal. along with a performance criterion. The theory then reveals an optimal algorithm which results in the minimum possible error, relative to the assigned performance criterion. The theory can also yield bounds on problem complexity (basically the cost to carry out the solution), and optimal information (of a given type and quantity). This latter, and newer, approach to dealing with problems involving uncertainty has been developed by J. Traub and co-workers and presented in the monographs [2,3]. More recently, it is begining to be discovered by engineers [5].

We conclude this section with a few remarks about the epistemological aspects of signal/data processing. In this, as in any other instance of scientific inquiry, it must be recognized that it is not possible to 'know reality' but, at best, how reality interacts with some sort of probe. That is, to paraphase Kant, our representation of things does not conform to these things as they are 'in themselves;' rather, they conform to our mode of representation. So we can make sense of the world only by imposing some structure originating from the mind upon it. This is the sense of the dictum of Protagoras that 'man is the measure of all things.' Therefore, we must acknowledge that the observer's presence is inevitable and ubiquitous in the final result because of the plan of observation. It follows that knowledge does not represent certainty, hence is not final, but rather open to improvement; we approximate truth by stages. We can never reach complete knowledge of reality, but we can obtain encodings (models) of it in terms of prior knowledge of the plan of observation. These models should be, as already suggested, the solutions to 'good questions;' each of them provides some insight into the overall situation, and a set of such models will eventually permit decisions to be made. We accept the unfortunate fate that most (all?) important decisions must be made on the basis of insufficient information, and we do as well as possible by means of the scientific method in general and data processing in particular.

Such philosophical speculations associated with scientific induction can be traced back to Plato and the question: "how can you seek what you do not know?". They were developed further by R. Descartes, I. Kant, D. Hume, and more recently by A. Eddington, K. Popper, R. Von Mises, *inter alia*. A convenient survey is given by W. Salmon in [6].

I.2 MATHEMATICAL METHODS IN SIGNAL PROCESSING

One way to acquire at least a superficial understanding of a scientific field is to organize it conceptionally into problems, methods, and results. We have already listed the major problems in signal processing in a generic way, and we will look at a few of these in more detail in Chapter III. We next want to briefly survey the major mathematical fields useful in signal processing, and then in Section I.3 a few examples of specific mathematical results that are of particular usefulness for signal processing. Several further mathematical models and theorems occur in the next chapter.

Once more we take time out to emphasize the subjective nature of our approach, as well as the eclectic nature of our subject. Signal processing draws upon many disciplines besides mathematics; for example, physics, systems and information theory, numerical analysis and computer science, digital circuit technology, etc. One need not be expert in all these fields (if it were possible!) in order to do signal processing. In particular, while a great deal of mathematics can be involved (as in physics), it is not all necessary to work productively on many problems. Thus each individual makes his own decision about how far to go in certain directions.

Having reiterated this position we now survey some useful areas of mathematics. As noted below in Section II.1, most practical data processing involves the manipulation of finite arrays of real or complex numbers. Thus it is immediately apparent that the methods of linear algebra will be crucial. Note only must the basic concepts [rank, (pseudo) inverse, eigenvalue/vector, condition, etc.] be mastered, but of equal importance is the need for stable numerical algorithms to compute these quantities. For example, a variant of Pisarenko's method of harmonic retrieval identifies the number of pure tones observed in noise as the rank of a certain hermitian matrix. For numerical determination of rank, pseudoinverse, and the solution of ill-conditioned systems of linear equations, the singular value decomposition is now the method of choice [7, 8]. Unfortunately, it is very computationally intensive and hence more suited to off-line treatment in sophisticated hardware. For real-time on-board applications the search of efficient high speed realizations of linear algebraic procedures continues, both in the areas of improved ('parallel') algorithms and a specialized hardware (array processors) utilizing VLSI technology.

Due to the ubiquitous presence of noise and impurities in observational devices, the data to be analyzed is never exact; that is, in the words of John Tukey, "what is measured is not the truth". Thus, as in many other aspects of life in general, it is necessary to look through noise when constructing estimates of unobservable signals. The mathematical methods for treating errors in data and the resulting estimates come from the fields of probability and statistics. The vital concepts are conditional distributions and expectations, limit theorems, the behavior of empirical distributions, probability density estimation, Neyman/Pearson theory, etc. Very commonly, received data is modeled as a realization of a stochastic process over some domain. The major theme of this report is to work with this setup under the further assumption that the domain has the structure of a *group*. Then the process can be viewed as a probability measure over a space of functions (sample paths) defined on the group. Such spaces often possess a great deal of rich mathematical structure (a Hilbert space, a Banach algebra, etc.), and so detailed (orthogonal) decompositions of the data are available.

We think it fair to say that the mathematical mainstream of the first half of the 20th century involved the simultaneous development of the manifestions of the concepts of groups and Hilbert spaces. These interlocked through the concept of a group representation, that is a continuous homomorphism of a given locally compact group into the unitary group on some Hilbert space. These developments were driven especially by the (then) new ideas and requirements of quantum mechanics, and are associated with the names of Hilbert, Weyl, von Neumann, Stone, along with, of course, many others. Other early motivations came from studies in integral equations (Riesz), formalization of Brownian motion (Wiener), models for prediction of time series (Wold, Kolmogorov), and Fourier series expansions (see below).

While it is true that all Hilbert spaces of the same orthogonal dimension are abstractly equivalent, they can differ greatly according to the nature of their elements. For signal processing applications useful Hilbert spaces can be considered to belong to one of three main types: L^2 spaces, reproducing kernel spaces (RKHS), and spaces of Hilbert-Schmidt operators acting on a fixed underlying Hilbert space. If P is a probability measure, $L^2(P)$ is the corresponding space of random variables with finite variance. If G is a finite or compact topological group, $L^2(G)$ is defined with respect to the associated Haar measure and forms a Hilbert space of exceptionally rich structure known as an H*-algebra [9]. The possibility of using such a space as a setting for signal or data models is provocative an is given a preliminary look in Chapter III. By contrast, reproducing kernel spaces generally contain very smooth or even analytic functions, and these elements can naturally serve as models of unobservable signals about which some information is available as data or constraints. Examples are the various Sobolev spaces whose elements are real-valued functions of one or several variables, each of a certain fixed degree of differentiability, Fock spaces of (Volterra) series of homogenious polynomials on a separable Hilbert space, and Hardy, Bergman, and Paley-Wiener spaces of analytic functions. The latter are a special importance, constituting as they do models of strictly bandlimited signals. Finally, Hilbert-Schmidt operators generalize matrices normed by the rule

$$\|[a_{ij}]\|^2 = \sum |a_{ij}|^2 \quad ,$$

which can serve as models of 2-dimensional data. On an L^2 space, for instance, the Hilbert-Schmidt operators are just the integral operators defined by a square integrable kernel. In general these operators are compact with square-summable singular values.

Now Hilbert space theory primarily concerns the study of operators acting between these spaces. In data processing operators occur both in the preliminary modeling and in the solution procedure. Thus we often assume that noise-free data occurs as the value assumed by some operator at the unknown signal. The operator models the effect of a communications channel and/or a measuring device, perhaps after some linearization and approximation. For example, a satellite detector might measure upwelling radiation at selected frequencies. This radiation at frequency ν is (approximately) related to temperature T by an integral operator of the form

$$I_\nu(p) = \int_{\Omega_p} K_\nu(p, q)T(q)dq \quad ,$$

where Ω_p is the volume within the detector field of view when the subsatellite point is p, and the kernel K_ν is determined from the equations of radiation transfer. Here we think of the atmospheric temperature T as the not-directly-observable signal of interest (perhaps as an input to some weather forecasting program), and values of I_ν as the data.

On the other hand, with the exception of some simple detection problems, virtually all data processing applications involve the transformation of one (received) signal into another. Hence by restricting attention to linear transformations and assuming that the data can be considered mathematically as belonging to some Hilbert space, a powerful and general theory can be built up from the already available operator theory. Let us therefore agree to define a (*linear*) *data processor* to be an operator acting from the data or sample space into a signal or model space. Note that once these two spaces are specified, then so are their operators. That is, the operators exist independently of any particular data analysis situation, and our task is to select one of them that best fits the available information and performance criterion. In practice, a suboptimal choice may be made for reasons of computational efficiency. This theme is elaborated on in Chapter III (see also the comments on the 'spline algorithm' in Section I.3).

At this point we have offered some general motivation for the use of linear algebra/operator theory and probability/statistics for signal processing models. We also alluded to the close connection between Hilbert spaces and groups through the representation concept. We now want to conclude this section with an attempt to delineate the basic role of group theory in signal processing. More detailed discussions of specific technical issues occur throughout the rest of this report.

It is commonplace that Fourier techniques (series and transforms) have been, and continue to be, of decisive importance in signal processing modeling. The historical reasons for this go back to the study of lumped linear time-invariant electrical circuits. The linear differential equations relating circuit voltage (or current) to external voltage (or current source) preserve the frequency of a sinusoided input. More generally, any linear time-invariant system has the harmonic exponentials $t \rightarrow e^{in\lambda t}$ as eigenfunctions, for appropriate choice of λ. Hence the system response to any (reasonable) input can be determined from the Fourier series expansion of that input. Fourier methods also permit solution of models for general electromagnetic radiation based on the wave equation. More recently, there has been the explosive development in computer-based methods for processing discrete data, by means of various 'fast' algorithms for the discrete Fourier transform. Sometimes the intent is to simply do fast digital data filtering, which is essentially a convolution of a data vector with the filter impulse response. Alternatively, spectral analysis of stationary data may be the goal, and here Fourier methods are required also of the very definition of the desired quantity (power spectral density function).

Now, all the Fourier transforms and expansions just alluded to are merely instances of a much more general situation. This is the field of 'abstract harmonic analysis', the study of functions defined on locally compact groups in terms of the associated (unitary) representations. We can't seriously contemplate summarizing this field which has been under active development since the publication of Weil's book [10] in 1941. The first English source for this material is Loomis [9], followed by Rudin [11] and the encyclopediac treatise of Hewitt and Ross [12], among others. However, we will offer just a few remarks aimed at providing a little perspective.

If we think of an element x of a (separable) Hilbert space as a mathematical model of a signal or a wave, we know from the elementary theory of such spaces that x can be expressed as a convergent series in terms of an arbitrary orthonormal basis $\{\ell_n\}$:

$$x = \sum c_n \ell_n,$$

and in fact the coefficients c_n are specified uniquely by the inner products $c_n = <x, \ell_n>$. A problem arises in view of the arbitrariness here: there is generally no natural way to carry out this decomposition. If, for example, x is a function defined on a compact interval then permissible choices of $\{\ell_n\}$ range from scaled trigonometric functions to orthogonal polynomials to step functions. To attempt a physical analogy we might say that, unlike a specific device (a clock, a car, etc.), a wave has no intrinsic parts.

A partial remedy exists if there is some additional structure available. Suppose in particular that our Hilbert space is $L^2(G)$, where G is some unimodular locally compact group and the integration is done with respect to its Haar measure. (The existence of a positive regular Borel measure on a locally compact group, which is invariant with respect to left, or right, translations, and which is unique up to a constant positive multiple, was the first major success of abstract harmonic analysis in the 1930s. Such measures are called left, or right, Haar measures, and permit the notion of invariant integration over G. Groups for which every left Haar measure is also a right Haar measure are termed unimodular. These include abelian, discrete, and compact topological groups, as well as semisimple Lie groups; in such cases we may speak simply of the Haar measure, which is unique up to a normalization constant. This latter is usually chosen so that the measure of G is 1 when G is compact, and so that the measure of each element of G is 1 when G is discrete, although this is clearly inconsistent when G is finite. When G is the additive group R^n, the Haar measure is proportional to ordinary Lebesque measure.) Suppose in particular that G is compact. Then there are distinguished orthonormal bases for $L^2(G)$, whose elements are of the form

$$\ell(g) = < U(g)y, z > , g \in G$$

where U is an irreducible unitary representation of G on a (necessarily finite dimensional) Hilbert space. If G is separable (an inessential restriction for practical purposes), there are only countably many inequivalent irreducible representations. The corresponding expansion of elements of $L^2(G)$ in terms of these special basis elements is a generalization of the classical Fourier series, to which it reduces when G = the circle group (the multiplicative group of complex numbers of modulus 1). Actually the classical theory is a bit easier because the circle group is abelian, so that the irreducible representations are one-dimensional (called 'characters' in this case), and hence can effectively be avoided. The basic point here is that the role of the simple harmonic exponentials $t \rightarrow e^{int}$ for the circle group is played more generally for any compact groups by its irreducible representations. This collection of results is known as the Peter-Weyl theory.

A final remark pertains to the all-important concept of the Fourier transform. This operator is defined for every $f \epsilon L^1(G)$ by the rule

$$\hat{f}(\gamma) = \int_G f(g) < \overline{g, \gamma} > dm_G(g) \tag{I.1}$$

if G is abelian and γ is a character on G, or by

$$\hat{f}(\lambda) = \int_G f(g)\, \lambda(g^{-1})\, dm_G(g) \tag{I.2}$$

if G is compact. In each case $m_G(\cdot)$ denotes the Haar measure on G. In Equation (I.1) $\hat{f}$ is a continuous function defined on the dual group Γ, which consists of all the (continuous) characters on G, while in Equation (I.2) $\hat{f}$ is defined on the unitary dual object Γ, which consists of all equivalence classes of (continuous) irreducible unitary representations of G. Note that in this latter case each value $\hat{f}(\lambda)$ is an operator on a certain finite dimensional Hilbert space. In each case $\hat{f}$ is termed the Fourier or *group transform* of f, and uniquely determines f. In the abelian case, there is a choice of Haar measure m on Γ so that if $\hat{f} \epsilon L^1(\Gamma)$ then the inversion formula

$$f(g) = \int_\Gamma < g, \gamma > \hat{f}(\gamma)\, dm_\Gamma(\gamma) \tag{I.3}$$

holds a.e. on G. (Usually, in fact, f is contrained to some class of smooth functions — the Schwartz class, the continuous positive definite functions, etc. — in such cases Equation (I.3) is valid on all of G.) Also in this case the transform can be defined on $L^2(G)$ so as to be a unitary with range $L^2(\Gamma)$. In the compact case, with proper interpretation of $L^2(\Gamma)$, the transform is again a unitary operator. These statements are known as Plancherel's theorem, and can, in fact, be extended to the general (separable) unimodular group [13], but this generality is not required below.

Since many of the classical groups are abelian the abstract group (Fourier) transform defined by Equation (I.1) extends and unifies all variants of the Fourier transform that occur in signal processing. All the familiar and important properties of the classical transforms remain valid. Thus, with the proper definitions, translations on G are converted into multiplications, and the convolution of two functions in $L^1(G)$ has a transform equal to the product of the individual transforms. Also, in the abelian case, there are a variety of results centered around the duality formula

$$(G/H)\hat{} = H^\perp \tag{I.4}$$

where H is a closed subgroup of G, and its annihilator $H^\perp = \left\{ \gamma \epsilon G : <h, \gamma> = 1,\ h \epsilon H \right\}$. For instance, the Haar measure $m_{G\ H}$ on the quotient group G/H may be chosen so that

$$\int_G f(g)dm_G(g) = \int_{G/H} dm_{G/H}(\xi) \int_H f(g+h)dm_H(h) \quad , \qquad (I.5)$$

a result known as Weil's formula. In Equation (I.5) m_H is Haar measure on H and ξ is the coset g + H.

We will be interested in the following generic applications of the Fourier transform. First, it serves to diagonalize convolution operators on a locally compact abelian or compact group. Such operators are discussed briefly in Section II.3 as generalizations of time-invariant filters, and in Chapter III as group filters. Second, on the finite groups of Chapter III there are fast algorithms for computing the Fourier transform. (Just how 'fast' such algorithms are for a particular group G, depends in an interesting way on the subgroup structure of G; the determination of this structure is a very difficult problem in general, but tractible in special cases. For instance, if G is abelian the issue reduces to the factorization of G into cyclic subgroups or, in an another approach, to the behavior of the group characters on the cosets determined by the members of a composition series. These fast transforms can in turn be used to compute group filters which, in turn, can serve as suboptimal approximants to Weiner filters. Further discussion is given in Chapter III.) Finally, partial Fourier transforms are often used for purposes of data compression and feature extraction for pattern recognition. We use this term to mean any linear transformation T on $L^2(G)$, where G is a unimodular locally compact group, of the form given by the right hand side of Equation (I.2), and where λ is any unitary representation of G. Thus if f is a received datum which can be considered to belong to $L^2(G)$, then T(f) is a statistic whose value, often called a *spectral component* or a *feature* of f, contains information about the underlying signal. Thus T(f) could be used as a basis for classifying the datum f into two or more pattern classes, or, if the dimension of the representation λ is small relative to that of $L^2(G)$, as one means of data compression. Note that for the practical case of finite dimensional data we have many choices for both G and λ, so this approach subsumes many special cases.

I.3 MATHEMATICAL RESULTS IN SIGNAL PROCESSING

Earlier we suggested the decomposition of a scientific field into problems, methods, and results, for the purpose of obtaining some insight into its activities. At this point we have discussed some problems and methods, at a fairly abstract level, and from a mathematical direction. We will conclude this chapter by indicating a few specific results.

Now evidently the huge literature in signal processing is teeming with 'results', so we are going to have to be rather choosy here. Before presenting our brief list of results we therefore indicate the criteria for their inclusion. First the result should be a definite, crisp, and nontrivial mathematical theorem, or at least a tightly knit collection of such. Further it should not just sit in solitary splendor but in fact it should have engendered significant further developments. Finally it, and/or some of these ensuring developments, should be widely used in signal processing practice. We also acknowledge the subjective nature of both these criteria, and of the decision as to whether this or that result meets all of them. Such subjectivity is a recurrent theme of this report.

Here is our list of results, followed by some brief comments.

- Karhounen-Loeve expansion
- Spline algorithm
- Maximum entropy principle
- Sampling theorem
- Kolmogorov isomorphism
- Fast Fourier transform

Experienced readers will, no doubt, consider other results deserving of mention; examples might be some version of a stochastic filter (Wiener, Kalman, . . .), the Levinson-Durbin method for fast solution of Toeplitz linear systems, etc. We simply feel that at least one of our three criteria are left unsatisfied by other results.

The original K-L expansion (1947) represented a stochastic process defined on a compact interval, with continuous covariance function, as an infinite linear combination of orthonormal functions, with mean square convergence. This had the practical effect of 'coordinatizing' the process by the countable set of random coefficients which, most importantly, turn out to be uncorrelated if the basis functions are chosen to the eigenfunctions of the integral operator defined by the covariance function. Truncation of this infinite expansion results in a minimum mean square error, for a fixed number of terms, and also minimizes an entropy function. Thus the K-L expansion is in several respects an optimal way to decompose (the sample functions of) the given process. Applications to signal detection soon followed (1950).

Nowadays there are many generalizations. The simplest is to replace the original process by a second order probability measure on a Hilbert space and to expand a random vector in the eigenvectors of the associated covariance operator. This operator, being self-adjoint and nuclear, has indeed an orthonomal basis of eigenvectors. and the resulting expansion converges with probability one. When the Hilbert space is finite dimensional there are a variety of applications to pattern recognition and data compression, and here the K-L expansion serves as a benchmark for the performance of other suboptimal but faster data processors (see Chapter III and, for example, [14]).

A different direction of generalization is to the case of Gaussian measures on a general Banach space; for example, the space of continuous functions on a compact metric space. From such work we learn, *inter alia*, that the K-L expansion over an interval converges uniformly with probability one, at least in the Gaussian case. For one such result, unifying several earlier ones we refer to [15] where, in particular, the role of the RKHS associated to a given Gaussian measure is stressed.

As a familiar example consider classical Brownian motion on the interval [0, 1]. Its covariance function is min(s, t) for $0 \leqslant s, t \leqslant 1$ and consequently its K-L expansion is a random Fourier sine series with zero-mean independent Gaussian coefficients. On the other hand, its associated RKHS is the Sobolev space consisting of absolutely continuous functions on [0, 1], vanishing at 0, and with square-integrable derivative. Proceeding in this direction we would be led to the classical Wiener measure on the Banach space of continuous functions C[0, 1], as the space of sample paths of Brownian motion, and then to the more recent unifying theory of abstract Wiener spaces. But that is another story [16].

The spline algorithm, as we are using this term, is a very general procedure for estimating an unknown element in a Hilbert space from partial information about it. The desired element is often construed as a model of some observed phenomenon, but the choice of the underlying Hilbert space is also part of the modeling procedure. In the simplest (noise-free) case it is assumed that a finite amount of linear information about the element is available, along with a bound on its norm. The optimal estimate, in a minimax sense, is then the value of the pseudoinverse of the data operator at the data vector. This is the classic prototype of a linear data processor as defined in the preceeding section. The optimal estimate, so obtained, is sometimes called an abstract interpolating spline. This is because when the data consists of a sampled values, and the Hilbert space is the Sobolev space of smooth functions on an interval, with square integrable second derivative, the estimate turns out to be the (unique) cubic spline interpolant of the data.

The early results on spline functions (piecewise polynomial functions joined smoothly, but not analytically, together) are due to Schoenberg and Sard in the late 1940s. The Hilbert space formalization, originally termed the 'hypercircle inequality' because of its geometric interpretation, was made by Golomb/Weinberger [17] and de Boor-Lynch [18]. During the last 15 years this basic result has evolved into two dynamic and powerful data processing methodologies: linear inverse theory (e.g., [19, 20]) and the theory of optimal algorithms [2] already mentioned in Section I.1.

One much-used application is the extrapolation of a bandlimited function from sampled data or, equivalently, the estimation of the spectrum (Fourier transform) from such data. In the latter case the resulting estimate has been termed the 'modified discrete Fourier transform' [21], and is used with over-sampled (higher than Nyquist rate) data. Another application occurs in the burgeoning field of tomography, where the reconstruction of cross-sectional tissue densities is attempted, based on the observed attenuation of a finite number of x-ray beams [22, 23].

The Principle of Maximum Entropy (PME) has a complex and controversial history, which is reviewed by E. Jaynes in [24], for example. Involved have been pioneers in the foundations of probability, from Laplace to Jeffries, of statistical mechanics, including Boltzmann and Gibbs, and of information theory, especially Shannon. In one direction, popularized by Jaynes and S. Kullback, it provides a systematic way of estimating probability distributions from known constraints; often these latter are certain moments of the distribution. Indeed, virtually all standard probability distributions can be so derived and characterized. PME has been used in statistical decision making to assign probabilities to possible outcomes, and therefore to permit business and economic decisions. Currently PME has been subsumed by PMCE, the Principal of Minimum Cross-entropy, a very general method of inductive inference wherein a probability distribution is singled out as 'closest' to a given prior from within the class of all distributions obeying known constraints. Here closeness of a pair of distributions p, q is measured by their cross-entropy

$$H(p, q) = \int \log(dp/dq)\, dp, \qquad (I.6)$$

assuming that p is absolutely continuous wrt q (if not, H(p, q) = + ∞). Among its many consequences PMCE can be used to neatly derive the classical method of maximum liklihood for statistical parameter estimation See [25] for a summary of the methodology, references to earlier work and applications to pattern classification, speech processing, image enhancement, and particularly to spectrum estimation.

It is this latter area, of course, that is of greatest overall significance in signal processing. Initially PME was introduced to the signal processing community by J. Burg in 1967 [26], as a new procedure for estimating the power spectrum of a stationary time series from partial knowledge of its autocorrelation function. Earlier approaches implicitly assumed this function to vanish for sufficiently large time lags, as they utilized the Fourier transform of the product of a window function of compact support with the known or estimated autocorrelation function. Burg's idea was to view the spectrum estimation as an infinite dimensional optimization, wherein the entropy of the process, taken proportional to the integral of the logarithm of the spectral density function (the underlying random process being assumed Gaussian), was maximized subject to the linear constraints imposed by the known values of the autocorrelation function. The solution turned out to be the spectrum of an autoregressive process of an order equal to the number of constraints less 1. In time, a better mathematical result has emerged, one that removes both the stationarity and Gaussian assumptions [27].

Nowadays the maximum entropy method is seen as the first of several parametric methods for spectral estimation, each involving a model fit, in some sense, to given time series data. These often provide superior frequency resolution compared with classical techniques. Meanwhile research based on PME has swung in the direction of multivariate spectrum estimation (see, e.g., the survey [28]), where, even with uniform sampling, the maximum entropy spectral estimate is not the same as an autoregressive model fit.

The remaining three of our distinguished results have some definite group-theoretic content and are consequently discussed in subsequent sections of this report.

REFERENCES

1. R. Holmes, "Mathematical Foundations of Signal Processing," SIAM Rev. **21**, 361-388 (1979).

2. J. Traub and H. Wozniakowski, *A General Theory of Optimal Algorithms* (Academic Press, New York, 1980).

3. ―――――― and G. Wasilkowski, *Information, Uncertainty, Complexity* (Addison-Wesley, Reading, 1983).

4. T. Kailath, ed., *Modern Signal Processing* (Hemisphere, Washington, 1985).

5. M. Milanese and R. Tempo, "Optimal Algorithms Theory for Robust Estimation and Prediction," **IEEE AC-30**, 730-738 (1985).

6. W. Salmon, *The Foundations of Scientific Interference* (University of Pittsburgh Press, Pittsburgh, 1966).

7. V. Klema and A. Laub, "The Singular Value Decomposition: Its Computation and Some Applications," **IEEE AC-25**, 164-176 (1980).

8. G. Golub and C. Van Loan, *Matrix Computations* (The Johns Hopkins University Press, Baltimore, 1983).

9. L. Loomis, An Introduction to Abstract Harmonic Analysis (D. Van Nostrand Co., Princeton, 1953).

10. A. Weil, *L'integration dans les Groupes Topologiques et Ses Applications* (Hermann et Cie, Paris, 1941).

11. W. Rudin, *Fourier Analysis on Groups* (Wiley, New York, 1962).

12. E. Hewitt and K. Ross, *Abstract Harmonic Analysis I and II* (Springer-Verlag, New York, 1963 and 1970).

13. I. Segal, "An Extension of Plancherel's Formula to Separable Unimodular Groups," Ann. Math. **52**, 272-292 (1950).

14. N. Ahmed and K. Rao, *Orthogonal Transforms for Digital Signal Processing* (Springer-Verlag, New York, 1975).

15. R. LePage, "Note Relating Bochner Integrals and Reproducing Kernels to Series Expansion on a Gaussian Banach Space," Proc. Am. Math. Soc. **32**, 285-288 (1975).

16. H. Kuo, *Gaussian Measures in Banach Spaces* (Springer-Verlag, New York, 1975).

17. M. Golomb and H. Weinberger, *Optimal Approximation and Error Bounds*, in Proc. Symp. on Numerical Approximation, R. Langer, ed., 117-190 (Univ. of Wisconsin Press, Madison, Wisconsin 1959).

18. C. de Boor and R. Lynch, "On Splines and Their Minimum Properties," J. Math. Mech. **15**, 953-970 (1966).

19. D. Oldenburg, "An Introduction to Linear Inverse Theory," **IEEE GE-22**, 665-674 (1984).

20. M. Bertero, C. de Mol, and E. Pike, "Linear Inverse Problems with Discrete Data, I: General Formulation and Singular Value Analysis," Inverse Problems **1**, 301-330 (1985).

21. C. Byrne and R. Fitzgerald, "Extrapolation of Bandlimited Signals: A Tutorial," in *Signal Processing: Theories and Applications,* M. Kunt and F. de Coulon, eds., 175-180 (North Holland, Amsterdam, 1980).

22. ————————, "Reconstruction from Partial Information with Applications to Tomography," SIAM J. Appl. Math. **42**, 933-940 (1982).

23. Special Issue on Computerized Tomograpny, Proc. IEEE **71** (March 1983).

24. *The Maximum Entrophy Formalism*, R. Levine and M. Tribus, eds., (MIT Press, Cambridge, 1979).

25. J. Shore, "Inversion as Logical Inference — Theory and Applications of Maximum Entropy and Minimum Cross-Entropy," in *Inverse Problems*, D. McLaughlin, ed., 139-149, SIAM-AMS Proc. **14** (1984).

26. J. Burg, "Maximum Entropy Spectral Analysis," at 37th Annual Meeting of Exploration Geophysicists, Oklahoma City, 1967.

27. B. Choi and T. Cover, "An Information-Theoretic Proof of Burg's Maximum Entropy Spectrum," Proc. IEEE **72**, 1094-1095 (1984).

28. J. McClellan, "Multidimensional Spectral Estimation," Proc. IEEE **70**, 1029-1039 (1982).

II. ASPECTS OF GROUP THEORY IN SIGNAL MODELING AND SAMPLING

Guided by the philosophy explained in the preceding chapter, namely, to search for mathematical paradigms and to justify operations on data, we review next some fundamental models of data processing with main emphasis on those where group theory plays a role. We stress that the attempt here is to describe a unified and systematic approach to a great diversity of problems. Hence a recurrent theme will be that even if a group is not immediately apparent in certain situations, it may be useful, in the above sense, to try to uncover a group 'lurking' in the background, or even to impose a group structure, on account of the immense body of theory and technique that then becomes available.

II.1 BASIC ALGORITHMS OF DATA PROCESSING

We begin with some empirical observations concerning the practice of computerized data processing. Whatever the original nature of the data, it is usually subjected to a series of pre-processing steps that serve to reduce it to finite dimensional form. Among these steps might be truncation, discretization and sampling, quantization, etc. This reduced data is called a block and its dimension the blocklength. The latter is determined by various factors, especially computer storage limitations, and the physical and statistical nature of the original data. In particular, successive data blocks must be treated as independent of one another, and only statistical associations between components of a block can be considered.

We also observe that most data processing algorithms involve a transformation of a data block into another block, possibly of a different length. Further, these transformations are usually linear, perhaps achieved as a composite of several relatively simple linear transformations. This is not surprising given the highly developed theory of linear transformations *vis-à-vis* any other class of transformations. We will continue this tradition and discuss only linear data processors.

Next, we observe that the most common data processing algorithms are (variants of) the Wiener-Kalman filter and the fast Fourier transform (FFT). In order to better focus our attention we will continue to enforce the assumption just made about the way the data is presented for processing, namely, block by block. This will eliminate from further consideration the Kalman-type recursive filters.

The term 'Wiener filter' is used generically for a linear transformation of the data chosen so as to minimize the average error in estimating a signal contained in the data. In the present simple situation of nonrecursive block data we can write

$$\underset{\text{(data)}}{y} = \underset{\text{(signal)}}{s} + \underset{\text{(noise)}}{\eta} \tag{II.1}$$

and then the Wiener filter W is defined by

$$E \| s - Wy \|^2 = \min \quad . \tag{II.2}$$

This minimization can be carried out by standard quadratic optimization techniques in the space of suitably dimensioned matrices, with inner product defined by <A, B> = trace (AB*); the result, for zero-mean signal and noise, is

$$W = C_{sy} C_y^{-1} = C_s (C_s + C_\eta)^{-1} \quad . \tag{II.3}$$

The middle term in Equation (II.3) is the product of the cross-covariance matrix of the signal and the data with the inverse of the covariance matrix of the data. Under the conventional assumption of signal and noise independence (or just zero-correlation), we have further $C_y = C_s + C_\eta$, as indicated.

A major extension of the foregoing model, deserving of brief mention here, is to the situation where the unknown signal x belongs to a Hilbert space H_1, is transformed by a linear operator A into a second Hilbert space H_2, and is observed there in the presence of noise process η, modeled as a zero-mean weak H-valued random variable. Thus

$$y = A(x) + \eta \quad . \tag{II.4}$$

The operator a represents the effect of a measurement device (probe) and/or a communication channel. Problems leading to such models abound in optics, geophysics, biomedicine, etc., where typically x is a function representing some physical variable of interest across a continuum of values of one or more variables.

Successful estimates of x in Equation (II.4) by means of a linear operator $B{:}H_2 \rightarrow H_1$ will depend on proper incorporation of prior information about x, either deterministic or stochastic. Usually the operator A is compact (if not actually of finite rank) and then methods involving pseudoinversions and singular function expansions can be employed. The concept of regularization, to compensate for the ill-posed nature of Equation (II.4), is important here. But all this is essentially pure Hilbert space theory and, so far at least, does not seem to have benefited from group theoretic techniques. So we will conclude this brief excursion by noting that the problem of recovering x from y in Equation (II.4), given prior information or constraints on x, is what we mean by a 'linear inverse problem', and that selected references have been provided following the discussion of the spline algorithm in Section I.3.

Returning now to our theme of basic algorithms, we next discuss the FFT. This is, of course, just an accelerated procedure for computing the discrete Fourier transform (DFT). The latter amounts to multiplying a given data block y by a particular unitary matrix F_N, where N is the blocklength of y and

$$F_N = w_N{}^{mn} \quad , \qquad 1 \leqslant m, n \leqslant N \quad ,$$

with $w_N = \exp(-2\pi i/N)$. Thus, the DFT is another linear transformation which can be applied to data vectors of arbitrary (finite) dimension. However, unlike the Wiener filter above, which is defined by a clearly stated purpose, it is not clear, *a priori*, why a DFT would be applied as part of a data analysis procedure. Indeed, the DFT is defined independently of any assumptions concerning the nature of the data. In fact, as we shall discuss in greater detail in the next chapter, the DFT is but one of a large class of unitary transforms associated with groups of finite order. These are the so-called group (Fourier) transforms, the formal definition of which was given in Section I.2 for locally compact abelian or general compact groups.

Also in the next chapter we will carefully examine the rationale for taking unitary transforms of data. Roughly a unitary transform represents the data vector in a new coordinate system while preserving the essential information contained in the data. Depending on the particular goal of the data processing we may expect a judiciously chosen unitary transform to reveal hidden features of the data (as in pattern recognition or spectral analysis), or to result in more nearly uncorrelated coordinates for quantization or coding purposes. Although there is, for any signal, an optimal unitary transform that decorrelates the signal, namely the discrete Karhounen-Loeve transform (DKLT), its practical usefulness is limited by severe computational difficulties, as well as possible lack of knowledge of the true signal statistics. Thus other unitary transforms, that are both data independent and computationally efficient, may be considered as suboptimal alternatives.

In addition to the role just described, group transforms also serve as components of a class of linear transformations called group filters. These transformations, which may equivalently be described as group convolutions, are again data independent and may, depending on the internal structure of the underlying groups, be computationally efficient. Hence group filters offer the possibility of doing fast suboptimal Wiener filtering. And, since convolutions are defined on all groups, abelian or not, we see that any (finite) group can be used to define a family of data processing operations, the success of which is a function of the group structure, the signal statistics and, of course, the overall purpose of the processing.

The upshot of this section has been to suggest, in addition to the well recognized roles played by Hilbert space/operator theory and by probability/statistics in signal processing, a significant role also for group theory. This particular role, namely the use of group transforms and filters, as fast and convenient approximations to a variety of data processing tasks, will be further described in the next chapter. The remainder of this chapter is devoted to a brief resume of several other aspects of group theory in signal processing. Each of these topics deserves more attention than can be provided in the present report. Hence they will be introduced merely to buttress our theme that a group-theoretic viewpoint is a fruitful one for many reasons in the design of signal models and data processors.

II.2 STATIONARY SIGNALS

Random data is usually modeled as a realization of a stochastic process. Thus a general mathematical task is to define and study relatively simple processes whose realizations, or sample paths, can replicate observations. Now a stochastic process consists of a family of random variables, whose values may be real, complex, or higher dimensional vectors. The family is often thought of as indexed by an integer or real variable representing 'time', but other index sets are not uncommon. For example, indices may comprise a set of 2- or 3-dimensional points, which correspond to the geometric distribution of sensors, whose outputs are the data. In any event, it is important to distinguish between the random variables which together define the process and the realizations, considered as functions on the index set. The process may equally well be viewed as a probability distribution on various function spaces over the index set, any one of which may be called the 'sample space'. (In particular, in the computerized processing of the preceding section, the data can be considered as a random sample from a finite dimensional distribution.) Both these distinct views of a stochastic process lend themselves to applications of group theory and the corresponding harmonic analysis. For the next couple of sections we will emphasize the former view, and the switch back to the latter (sample space distribution) view.

Classically, stationary processes were proposed as models for signals whose statistical fluctuations appeared to be independent of time. The simplest of several possible precise definitions is that the process mean should be a constant, and that the covariance function evaluated at points s, t, should depend only on the difference t-s. Such processes are called weakly (or, wide-sense) stationary. More restrictive definitions of stationarity may be given; these involve invariance to time shifts of higher moments, or of the entire family of finite dimensional distributions. However, these will not be required for what follows.

As far back as 1948 it was recognized that this definition involved the (additive) group structure of the integers or the real numbers, according as time was represented discretely or continuously. It was then a short leap (as least for mathematicians!) to extend the definition of weak stationarity to processes defined on any locally compact abelian (lca) group. Thus if G is such a group, and L_0^2 (P) is the space of zero-mean second order random variables with respect to the probability measure P, we say that the (continuous) mapping $g \rightarrow x_g$, from G into L_0^2 (P), is a *weakly stationary* stochastic process on G if

$$\begin{aligned} E\,(x_g \bar{x}_h) &= E\,(x_{g+k} \bar{x}_{h+k}) \\ &= E\,(x_{g-h} \bar{x}_e) \\ &= r(g-h) \end{aligned}$$

for all g, h, k ϵ G, and e = identity element of G; the function r just defined on G is the covariance function of the process. A final bit of abstraction can be obtained by replacing the space L_0^2 (P) by an arbitrary Hilbert space H and then defining the function r by

$$<(x_g, x_h> = <x_{g-h}, x_e> = r(g-h) \quad . \qquad (II.5)$$

It is remarkable that so much structure ensues from this simple hypothesis of weak stationarity. First of all, it turns out that this concept is fundamentally linked with that of a unitary representation of G in H. Indeed, given such a representation U, that is, a (continuous) homomorphism from G into the group of unitary operators on H, and any vector x ϵ H, then

$x_g = U(g)\cdot x$ is weakly stationary. And conversely, if x_g is a given weakly stationary mapping of G into H, with covariance function r, there exists a unitary representation U of G on H, and $x \in H$ such that $r(g) = \langle U(g)x, x\rangle$. These facts do not require that G be abelian, and will reappear in Section III.3 in the context of general finite groups.

Next, it follows from Bochner's theorem that, owing to the positive definiteness of r, there is a finite positive (regular Borel) measure μ on the dual group Γ such that

$$r(g) = \int_{\Gamma} \gamma(g)d\mu(\gamma), \quad g\epsilon G \quad . \tag{II.6}$$

This μ is called the spectral measure of the process $\{x_g\}$, and its Radon-Nikodym derivative with respect to the Haar measure on Γ is called the spectral density of $\{x_g\}$.

Equation (II.6), a general version of the so-called Wiener-Khinchine relation, connects a 'time-domain' concept, the covariance function r, with a 'frequency-domain' construct, the spectral measure μ. In particular, it exactly locates the proper domain of definition of the latter as the (Borel field of) the dual group. Thus given a weakly-stationary process defined on any group, discrete or not, of any dimension, we know where to set up a frequency domain analysis. In fact, we can achieve a very tight yet decoupled relation between these two domains, as indicated next, using another group theoretic result.

Given a weakly stationary H-valued process $\{x_g\}$, $g\epsilon G$, we can make the association

$$T{:}x_g \to \langle g, \cdot\rangle$$

between the element x_g in H and the character defined by g on Γ. Because the spectral measure μ is finite, these characters generate the space $L^2(\Gamma, \mu)$, and it turns out the T can be extended to an isomorphism between this space and the closed subspace H_x of H spanned by $\{x_g\}$. This result is known as the *Kolmogorov isomorphism*, since Kolmogorov orginally proved the special case where G is the group of integers. In this way we can replace the rather mysterious space H_x, in practice consisting of linear functions of the random variables $\{x_g\}$, by a more familiar function space $L^2(\Gamma, \mu)$.

The isomorphism just mentioned can be explicitly implemented in terms of a stochastic integral, which can in turn be rather easily derived from a generalized version of Stone's theorem. This generalization states that any (weakly) continuous unitary representation U of an lca group G can be expressed as an integral of a 'resolution of the identity' E on Γ:

$$U(g) = \int_{\Gamma} \langle g, \gamma\rangle \, dE(\gamma), \quad g\epsilon G \quad . \tag{II.7}$$

Here E is a strongly countably additive measure on Γ whose values are orthogonal projections on H, and with $E(\Gamma) = I$. Thus U is a kind of abstract Fourier transform of the projection-valued measure E. The validity of Equation (II.7) proceeds from Bochner's theorem and some general measure theory.

It follows that we can express our original H-valued process $\{x_g\}$ in the form

$$x_g = U(g)x_e = \int_\Gamma <g, \gamma> dE(\gamma) \cdot x_e \quad ,$$

where now the integrator is the H-valued orthogonally scattered measure W whose value at any Borel set $B \subset \Gamma$ is the vector $W(B) = E(B) \cdot x_e$. When $H = L_0^2(P)$ this expression is the stochastic integral just mentioned, and defines the inverse of the Kolomogrov isomorphism:

$$T(x_g) = <g, \cdot> \quad ;$$
$$x_g = T^{-1}(<g, \cdot>) = \int_\Gamma <g, \gamma> dW(\gamma), \quad g \epsilon G \quad . \qquad (II.8)$$

Classically, this formula is Cramer's representation of the process as the Fourier transform of a random measure with orthogonal increments. Finally, the connection between Equations (II.6) and (II.8) is simple: $\mu = \| W(\cdot) \|^2$.

With these general principles established, more specialized models can be developed, based on the integration of various stochastic measures. For example, we may say that a 'white noise' on the group G is a stochastic measure W whose associated scalar measure $\mu_W(= \| W(\cdot) \|^2)$ is a Haar measure on G. The convolution of an L^2-function ϕ on G with W then results in a certain weakly stationary process

$$x_g = \int_G \phi(g - t)dW(t) \quad , \quad g \epsilon G \quad , \qquad (II.9)$$

which is a generalization of the classical 'moving aver ge'. These latter occur when G is the group of integers $Z = \{n\}$; the white noise reduces to a sequence of zero-mean uncorrelated random variables $\{W_n\}$ of variance 1, and ϕ becomes a square summable sequence $\{\phi_n\}$. Then Equation (II.9) is just

$$x_n = \sum_{k=-\infty}^{\infty} \lambda_{n-k} w_k \quad , \qquad (II.10)$$

the usual moving average representation.

In general, a weakly stationary process has the form of Equation (II.9) if and only if its spectral measure is absolutely continuous wrt Haar measure on Γ[39].

A more leisurely presentation of the preceding ideas is given in [1] for the special case of discrete processes, that is, the case where G is the group of integers. The theory of (Fourier) analysis on lca groups is succinctly set forth in [2, Chapter 1]. The general theory of orthogonally scattered measures is due to P. Massani [3]; a brief review occurs in [1]. Second order weakly stationary processes were first defined on lca groups by J. Kampé de Fériet; a more recent treatment is [4]. The survey paper [1] contains numerous further references.

I᾿.3 TIME-INVARIANT FILTERS

Filters are devices which purposefully modify data as, for instance, the Wiener filter already discussed in Section II.1. Thus, mathematically speaking, a filter can be viewed as simply a (usually linear) transformation of data, and indeed this is the general view taken in this report. But traditionally the term 'filter' is used in a more specific context, to denote a linear transformation which is time-invariant and causal (nonanticipating). Thus a filter appears as a special type of operator on a sample space for the underlying process whose realizations are the possible data. As such, the filter is defined somewhat independently of the process; the main requirement being that the domain of the filter contain the path space of the process.

Now w..en the sample paths are defined along the real line, or a discrete subset thereof, the filter is termed 'time-invariant' if a shift in an input to the filter is preserved in the output. This notion can be readily extended to the group context: if G is an lca group, and T a filter whose domain is a translation-invariant space M of functions defined on G, the T is called *invariant* if

$$T\,[u(\cdot - g)] = T(u)\,(\cdot - g) \quad , \tag{II.11}$$

for each $g \epsilon G$ and $u \epsilon M$. That is, T commutes with the family $\{\tau_g \cdot g \epsilon G\}$ of translation operators defined by the elements of G:$T \cdot \tau_g = \tau_g \cdot T$, where by definition $\tau_g(u)(x) = u(x - g)$, $g \epsilon G$, $u \epsilon M$.

The mathematical issues here inclu'le the following:

(a) For a given translation-invariant space M, what is the structure of operators T that satisfy Equation (II.11)? This question may be extended to include cases where the range of T lies in a second translation-invariant space.

(b) When does there exist a 'frequency-response function?' Based on the classical situation this should be a (measuraole) function ϕ, defined on the dual group, such that the action of T is equivalent under the group (Fourier) transform to multiplication by ϕ. *A fortiori*, the space M of (a) is now $L^2(G)$.

(c) When can the stochastic process $\{x_g : g \epsilon G\}$ which is generating our observations be realized in $L^2(G)$, or in some other translation-invariant subspace M for which the answer to the question of (a) is 'interesting'?

These are questions difficult to answer at a high level of generality. Most of the available answers (a) and (b) have, in fact, only become available in the last 20 years, primarily from research in harmonic analysis. Here we just indicate a couple of special cases. First let T be a continuous linear transformation either from $L^2(G)$ into $C(G)$, or from $L^1(G)$ into $L^2(G)$, and suppose that T commutes with all the translation operators τ_g [here C(G) is the space of all continuous functions defined on G] Then T is convolution with a fixed L^2 function ϕ; that is,

$$T(f)(x) = f * \phi(x) = \int_G f(x - g)\phi(g)dg \quad . \tag{II.12}$$

This was established in [5] along with several other similar results.

Let us next consider the case where T is an operator on $L^2(G)$ that commutes with translations. Here the situation is a bit murky, but the basic result is that any such T is convolution with a 'pseudomeasure' on G [5, 7]. This includes, but is not limited to, the case where there is a bounded measure μ on G for which

$$T(f)(x) = f * \mu(x) = \int_G f(x - g)\, d\mu(g) \quad .$$

Things become a bit clearer (as is so often the case!) by taking a group (Fourier) transform. Under its action, suitably defined [5, Section IV], pseudomeasures corresponding uniquely to (locally) essentially bounded functions on the dual group, and the operation of convolution with a pseudomeasure goes over into multiplication by the corresponding function. In this way we obtain a general answer to (b).

A similar result, but allowing T to be only closed and densely defined, later obtained in [6], by very different (operator-theoretic) methods, where it was termed a 'generalized Bochner theorem'. The original Bochner theorem (1929) pertained to the case where G is the group of real numbers. So, the upshot is that, when G is a group, any invariant filter on $L^2(G)$ is unitarily equivalent, via the group (Fourier) transform, to a bounded multiplication operator on $L^2(G)$.

In general, any operator acting between a pair of translation-invariant function spaces defined on an lca group G, and commuting with the translation operators, is termed a 'multiplier' for that pair. Thus, for example, it can be shown that the correspondence $T \rightarrow \phi$ defined in Equation (II.8) above is actually an isometric isomorphism between the space of all multipliers for the pair $[L^1(G), L^2(G)]$, and the space $L^2(G)$. Similarly, the space of multipliers of $L^2(G)$ is isometrically isomorphic to $L^\infty(G)$, and also to the (suitably defined and normed) space of pseudomeasures on G. Which form of the multiplier (invariant filter) we use depends on whether we want to operate in the 'time domain' G or the 'frequency domain' Γ. The abstract theory of multipliers is surveyed by Larsen [7].

Finally, as to question (c), we observe that many interesting processes on G cannot be realized in $L^2(G)$, although there is, of course, no problem when G is of finite order (the case of interest in the next chapter). Otherwise, we have a simple sufficient condition, under a mild measurability restruction on the process, that the function $g \rightarrow \mathrm{var}(x_g)$ be integrable. It is interesting to note (for some, if not for present purposes) that a (measurable) second order process on G can always be realized in some weighted L^2 space over G [8]. For this result the group structure of G plays no role and, indeed, such spaces need not be translation invariant.

In this brief review of sample space filtering we have concentrated on understanding the time-invariance aspect in the general group context. Causality depends on an ordering of the group, so that the terms 'past' and 'future' have a meaning. When G is a subgroup of the real numbers the frequency response function of a stable casual time-invariant filter can be extended to an analytic function. But it seems to be somewhat artificial to try to define these terms in general, so we will not pursue the matter further.

A second point oi view of the concept of a filter is to view such as a linear transformation on the (Hilbert) space generated by the random variables comprising the process of interest. Thus let $\{x_g:g\epsilon G\}$ be a weakly stationary H-valued process defined on the lca group G. As before, the subspace of H spanned by the x_g is denoted H_x. We know from Section II.2 that there is a unitary representation $g \rightarrow U(g)$ associated with this process. This in turn, according to Equation (II.7), is defined by a resolution of the identity E(·) on Γ. An operator T on H_x might now be called time-invariant (sometimes also called 'U-stationary') if it commutes with all U(g), $g\epsilon G$. A question of long interest is to determine the structure of such operators, which for short we may now refer to as 'filters'. The basic goal is to assert that T is an E-integral; that is,

$$T = \int_{\Gamma} \phi(\gamma) dE(\gamma) \quad , \qquad \text{(II.13)}$$

where ϕ is some measurable function on G. This equation means that for $y\epsilon H_x$,

$$T(y) = \int_{\Gamma} \phi(\gamma)\, dW_y(\gamma) \quad , \qquad \text{(II.14)}$$

where W_y is the orthogonally scattered H_x-valued measure defined by $W_y(B) = E(B) \cdot y$, $B \subset G$. Actually, the integral in Equation (II.14) can only be defined provided that $\phi \in L^2(\mu_y)$, where μ_y is the spectral measure on G associated with W_y: $\mu_y = \| W_y(\cdot) \|^2$. This requirement, sometimes termed the 'matching condition' in the signal processing literature [10], naturally holds when ϕ is (essentially) bounded. The function ϕ may now be termed the frequency response function of the filter T. Conditions for the validity of Equation (II.13) are provided in [6, 9]; they involve further restrictions on either the group $U(g):g\epsilon G$ or on the operator T. The simplest of these conditions [6] is that there should exist a cyclic vector in H_x for $U(g):g\epsilon G$ or, in other words, this group of operators should have unit multiplicity.

If $y_g = T(x_g)$, $g\epsilon G$, represents the output of the filter T, it is clear that $\{y_g\}$ is again weakly stationary with the same shift group $\{U(g):g\epsilon G\}$ as $\{x_g\}$. The value of the representation (II.13) for T is that it permits an easy but rigorous derivation of the 'frequency domain' behavior of the filter. Namely, if W_x and W_y are the orthogonally scattered measures corresponding to $\{x_g\}$ and y_g, respectively, then it follows that

$$W_y(B) = \int_{B} \phi(\gamma)\, dW_x(\gamma)$$

for all Borel sets $B \subset \Gamma$. For the associated spectral measures this implies that

$$\mu_y(B) = \int_{B} \phi(\gamma)\, dW_x(\gamma)$$

and consequently that $d\mu_y/d\mu_x = |\phi(\cdot)|^2$. Finally, we see that if μ_x is absolutely continuous wrt Haar measure on G, then so is μ_y; if so, the corresponding spectral density functions f_x and f_y are related by

$$f_y = |\phi(\cdot)|^2 f_x \, , \quad \text{ae} \quad .$$

This is a fundamental relationship in signal processing and points up the important role of the function $|\phi(\ )|^2$, known as the 'filter gain'.

Derivation of these results and discussion of some of their implications is given in [1]. We emphasize once more that we have indicated two complementary approaches to rigorous time-invariant filter definition and design over groups: filtering can be done by an operator either on the sample space or on the space generated by random variables. In the first case the filter is defined independently of the process; otherwise, it is defined directly in terms of the unitary shift group of the process, which is assumed weakly stationary.

II.4 NONSTAT!ONARY SIGNALS

Unfortunately, the elegant theory of weakly stationar signals fails to encompass many signals of interest. Such signals may exhibit a time-dependent mean (a 'trend') and other higher moments. There are two generic approaches to this real difficulty: transform the data to regain stationarity or recognize the data as belonging to a broader signal class for which some (useful) structure theory exists. The first approach includes such filtering tasks as detrending and general 'prewhitening' of the data along with statistical tests of stationarity applied to the residuals. The other approach is of greater mathematical substance and involves defining and characterizing larger classes of signals, hopefully retaining some of the useful results available in the stationary class.

This transition away from stationarity bears some analogy with similar movements in more familiar settings: from linear to nonlinear differential equations, say, or from normal to non-normal operators. In each of these contexts we leave behind a highly developed and successful discipline to encounter a comparative wilderness, which can at best be comprehended by a variety of special cases, techniques and approximations.

A succession of attempts by eminent probabilists (Loéve, Karhounen, Cramér, Bochner, Rozanov, Rao) to define a viable extension of (weak) stationarity began almost 40 years ago and has culminated in the concept of (weak) harmonizability. It is now understood that this concept may be approached in several equivalent ways, each resulting from a generalization of the corresponding construct in the stationarity theory. Thus one may attempt to extend the Weiner-Khinchine relation Equation (II.6) between the covariance function and the spectral measure of the process, the Cramer representation Equation (II.8) of the process as the Fourier transform of an orthogonally scattered vector measure, or the operator description obtained from Equation (II.7) and the concept of a unitary representation. In particular, the dependence of the covariance functions or equivalently, the spectral measure, on a single variable must be relaxed. Yet at the same time it is desirable to maintain ties with Fourier analysis so as to conserve the frequency interpretation of linear filtering.

This is not the place to delve into the many measure-theoretic technicalities required to precisely make the various definitions of weak harmonizability and to establish their equivalence. For such details the recent survey of M. Rao [11] may be consulted. Here we will just take note of a few highlights.

Probably the most straightforward definition is that an H-valued *weakly harmonizable* process on an lca group G is a bounded weakly continuous mapping $g \rightarrow x_g$ from G into H of the form

$$x_g = \int_{\Gamma} <g, \gamma> dW(\gamma) \quad , \qquad \text{(II.15)}$$

where W is a vector measure on the Borel algebra of the dual group Γ. Thus W is merely countably additive and no longer necessarily orthogonally scattered as in the earlier stationary case Equation (II.15) is a very general Fourier transform relation, so that this new concept conserves some link with harmonic analysis.

One immediate consequence of this definition is that if T is a bounded linear operator on H, and $\{x_g : g \epsilon G\}$ is a weakly harmonizable H-valued process, then so is $y_g = T(x_g)$. Thus this new class of processes is closed under the application of arbitrary linear operations, a situation which definitely does not obtain for the stationary processes (or even for other more restricted definitions of harmonizable processes). This fact has the implication that any well-defined linear filter, time-invariant or not, applied to a weakly harmonizable input yields an output of the same sort, an observation first made in a more restricted fashion in [12]. Hence, for systems analysis purposes this is as wide a class of processes as needs be considered.

A second, but less immediate consequence of the definition is the following expression for the covariance function R of the process:

$$R(g, h) = <x_g, x_h> = \int_\Gamma \int_\Gamma <\overline{g, \lambda}> F(d\gamma, d\lambda) \quad , \qquad (II.16)$$

where the integrator set function F is defined by

$$F(A, B) = <W(A), W(b)> \quad , \qquad (II.17)$$

for A, B, Borel subsets of Γ. The integration in Equation (II.16) can be a little tricky since F need not define a measure on $\Gamma \times \Gamma$, unless it is of bounded variation. However, when the proper integral is used (the so-called Morse-Transue integral) as explained in [11], then the covariance formula Equation (II.12) provides a characterization of weakly harmonizable processes. Even when F does define a *bona fide* measure, which we might then refer to as the spectral measure of the process, we observe from Equation (II.17) that it will generally be complex-valued, unlike the positive spectral measures corresponding to the weakly stationary processes. Weakly harmonizable processes with a spectral measure arє now commonly referred to as strongly harmonizable and are often encountered in the engineering literature (primarily for the case G = $\{$real numbers$\}$; [12, 13]).

If $\{x_g : g \epsilon G\}$ is a weakly stationary H-valued process, and P is an orthogonal projection on H, then by what has already been noted the process $\{P(x_g) : g \epsilon G\}$ is weakly harmonizable. Remarkably, there is a valid converse statement which provides an elegant characterization of weakly harmonizable processes [11, 14]. Namely, let $\{y_g : g \epsilon G\}$ be weakly harmonizable in the Hilbert space H. Then there exists a larger Hilbert space K and a weakly stationary K-valued process $\{x_g : g \epsilon G\}$ such that $y_g = P(x_g)$, $g \epsilon G$, where P is the orthogonal projection from K onto H. Thus each weakly harmonizable process on the lca group G appears as a projection of an associated weakly stationary process on G, defined in an enlarged Hilbert space. Equivalently, a weakly harmonizable process can always be 'dilated' to a weakly stationary process. This fact is naturally related to previously known results concerning unitary dilations of contraction operators, and to Naimark's theorem concerning the dilation of a positive-definite operator function on G to a unitary representation of G [15].

A rather different approach to the treatment on nonstationary signals also originated in the mid 1940s, and continues vigorously into the present time. This approach, the joint time-frequency representation of signals, emmanates primarily from the community of physicists and electrical engineers, and has associated with it the names of Wigner, Gabor, Ville, Woodward, Rihaczek, Cohen, *inter alia.* The analysis and design of radar waveforms [16, 17] was a primary engineering motivation. More recently, this work has been applied to the detection of phase-modulated signals in noise [18]. In a different direction efforts have been made to absorb some of the resulting constructs (such as the radar ambiguity function) into a general mathematical framework, specifically, that of nilpotent harmonic analysis [19, 20]. The key mathematical object in this work is a certain Lie group known as the real Heisenberg group.

The essential idea here is that the energy of nonstationary signals is distributed in both time and frequency. This is already clear with audio signals where both the pitch and the time of origin of a tone can be heard. It is also a well-known aspect of radar signals where the echo is subjected to both a time delay and a frequency (Doppler) shift, depending on the range and radial velocity of the target. Hence it is desirable to express the signal as a function of both time and frequency. Without trying to be overly detailed we next indicate two generic approaches to this problem.

As we know from Section II.2 the spectral content of a weakly stationary process is independent of the time index. In the standard cases of engineering practice, this index runs through either the group of real numbers or a discrete subgroup thereof. In such cases consistent estimates of the spectral density function of the process are conventionally obtained as the squared magnitude of the Fourier transform of a suitably windowed segment of a sample function of the process. (Naturally, an ergodic hypotheses must be invoked here for the validity of such inferences.) The frequency resolution of the resulting estimates then depends on the window length. Now when the signal is not stationary we may, on the one hand, try to select a window of sufficiently short length that the signal portion under this window is approximately stationary. In this way we are led to a compromise between resolution in time and in frequency. Sliding a given window along the data then permits a display of the variations in frequency as a function of time. Such joint functions of time and frequency are called 'spectrograms' [21].

A second approach to nonstationary signal analysis is to attempt a direct definition of a function of time t and frequency ω which somehow measures the distribution of signal energy over the (t, ω) plane. There are several desiderata here. The correspondence between a signal f and its proposed distribution F should ideally be

(1) bilinear in f

(2) non-negative

(3) possessed of correct marginals.

This last requirement refers to the result of integrating F over all t or over all ω:

$$\int F(t, \omega)\, d\omega = |f(t)|^2 \quad ,$$

$$\int F(t, \omega)\, dt = |\hat{f}(\omega)|^2 \quad ,$$

where $\hat{f}$ is the suitably normalized Fourier transform of f. Unfortunately it is not possible to satisfy all these three desiderata simultaneously.

At present there are many choices available of what might be called 'pseudodistributions', that is, functions F obeying some of the above criteria. For instance, the rule

$$\begin{aligned} F(t,\omega) = \frac{1}{4\pi^2}\int\int_{-\infty}^{\infty}\int e^{-i\theta t - i\tau\omega + i\theta u}\, k(\theta,\tau) \quad . \\ f(u+\tau/2)\,\bar{f}(u-\tau/2)\,du\,d\tau\,d\theta \quad , \end{aligned} \qquad \text{(II.18)}$$

where k is a rather arbitrary kernel subject to $k(0,\tau) = k(\theta,0) = 1$, has the correct marginals, and is bilinear in f provided that k is independent of the signal f. These distributions are said to constitute *Cohen's class* [22]. Special choices of k yield many popular distributions, including those of Wigner (take k = 1) and Rihaczek [take $k(\theta,\tau) = \cos(\theta\tau/2)$]. The Wigner distribution is of interest for several reasons: its 2D Fourier transform is the familiar ambiguity function of radar theory and any member of Cohen's class can be obtained from it by convolution with an appropriate measure. Further, in a certain technical sense, the Wigner distribution comes the closest from Cohen's class to being non-negative [22]. In addition to these somewhat theoretical reasons the Wigner distribution has been applied to a variety of practical data processing tasks; see [18] and the thesis [23] together with its references.

The study of the many properties, interrelations and applications of these functions is a major occupation of modern signal processing workers. Extention of these aspects from the real line to its discrete subgroups has begun, and in time we may expect the theory to slowly progress to signals defined over more general lca groups.

In addition to the aspect just suggested of one possible role of group theory in joint time-frequency signal analysis, there is another, both more profound and less expected. Restricting our attention for the remainder of this section to signals defined on the group R of real numbers, we are making reference to the appearance of the real three-dimensional Heisenberg group H. According to Schempp [19], this group "... stands at the crossroads of quantum mechanics and signal theory." Again, without intending to be overly detailed, we indicate a little of the relevant background.

The mathematical embodiment of the uncertainty principles of conjugate quantities in nonrelativistic quantum mechanics occurs both as Heisenberg's inequality for Fourier transforms of functions in $L^2(R)$:

$$\int_{-\infty}^{\infty} t^2|f(t)|^2\,dt \cdot \int_{-\infty}^{\infty} \omega^2|\hat{f}(\omega)|^2\,d\omega \geqslant E_f^2/16\pi^2 \qquad \text{(II.19)}$$

where $E_f = \|f\|_2^2$, and as the commutativity relation

$$PQ - QP = cI \tag{II.20}$$

for a corresponding pair of self-adjoint operators P, Q and a pure imaginary constant c. In the classical single particle case, P and Q are the position and momentum observables, and $c = h/2\pi i$, h = Planck's constant. The Heisenberg inequality then expresses the impossibility of exact simultaneous measurement of both these quantities. It does this by the interpretation of Equation (II.19) as a lower bound on the product of the variances of the observables P and Q in any state f. (When $E_f = 1$, the function

$$B \rightarrow \int_B |f(t)|^2 dt$$

defines a probability measure on the real line R, and is considered to define the probability that our particle is found in the Borel set B.)

The relation (II.20) was given a group theoretic interpretation long ago by Weyl, who replaced the operators P and Q by the one-parameter groups on unitary operators which they generate. Further, a connection was made with the Heisenberg group H_1 already mentioned. By definition, H_1 is the subgroup of all 3 × 3 real matrices of the form

$$g = \begin{bmatrix} 1 & a & c \\ 0 & 1 & b \\ 0 & 0 & 1 \end{bmatrix} \tag{II.21}$$

As such, H_1 is both noncompact and nonabelian. It turns out that there is a deep connection between the group version of Equation (II.20) and corresponding commutativity relations between the one-parameter subgroups of H obtained from the matrices Equation (II.21) by fixing two of a, b, c at zero. From this one can eventually obtain a description of the irreducible unitary representations (of dimension > 1) of H_1 on $L^2(R)$. These have the form

$$U_\lambda(g)f(t) = e^{i\lambda(c+ta)} f(t+b) \quad , \quad f \in L^2(R) \quad , \tag{II.22}$$

for $g \in H_1$ and for any real $\lambda \neq 0$. In particular, U_1 is often called the linear Schrödinger representation of H_1. For this background material we refer to [24, 25].

Now, even though there is no analogue of Planck's constant in signal theory, there is a well-known uncertainty principle that applies to radar measurements. It essentially places limits on achievable resolution performance in range and range rate. This principle can be arrived at through an analysis of the ambiguity function A. As already noted, A is the Fourier transform of the Wigner distribution W, a special case of Equation (II.18). Up to an inessential phase factor the ambiguity function is obtained by cross-correlating a given signal with its time and frequency shifted version:

$$A(\tau, \omega) = \int_{-\infty}^{\infty} f(t + \frac{\pi}{2}) f(t - \frac{\tau}{2}) e^{-i\omega t}\, dt \quad . \tag{II.23}$$

Unlike the real-valued W, this function may be complex-valued. Nevertheless, we always have

$$|A(t, \omega)| \leqslant A(0, 0) \quad ,$$

and

$$||A||_2 = A(0, 0) = E_f \quad ,$$

indicating a constant volume under the surface $|A(\tau, \omega)|$, independent of the signal f. This implies the impossibility of concentrating A (or W, for that matter) around a particular point, as would be desirable for separating closely spaced targets. In general, it follows that separability in one variable is only to be gained at the expense of self-clutter and masking in the other variable.

Now, rewriting the ambiguity function A in Equation (II.23) as

$$A(t, \omega) = \phi \int_{-\infty}^{\infty} f(u + \tau)\, \overline{f}(u)\, e^{-i\omega u}\, du \quad ,$$

($|\phi| = 1$) and comparing with the representation formula (II.22), we see that

$$A(t, \omega) = < U_1\,(g_0)f, f > \quad . \qquad \text{(II.24)}$$

Here g_0 is restricted to elements of the form (II.21) with c = 0 (and with the identifications $a = -\omega$, $b = \tau$), and the bracket notation on the right hand side refers to the inner product in $L^2(R)$. Intuitively, we think of f as the envelope of a radar pulse of finite energy and Equation (II.24) expresses the cross correlation of the pulse with its echo.

The foregoing relation (II.24) is the basic connection between the theory of radar waveform design and harmonic analysis on the real Heisenberg group. For a detailed survey of this link and its many consequences, one may consult the papers of Schempp [27 + cited literature], and [20]. Unfortunately, the mathematical prerequisites for a careful development of this material are rather severe.

The Heisenberg group concept is somewhat more general and ubiquitous than might be inferred from the preceding remarks. Given any commutative ring with identity or any lca group, an associated Heisenberg group can be defined. Thus, in the ring case, we can use the matrices of the form (II.21), with a, b, c ring elements. Or, in the case of an lca group G, we can take the set $G \times \Gamma \times T$ with multiplication

$$(g, \hat{g}, s) \cdot (h, \hat{h}, t) = (g + h, \hat{g} + h, st < g, \hat{h} >)$$

as the associated Heisenberg group; here T is the circle group. The resulting construct plays a key role in many aspects of the accompanying harmonic analysis [26-30].

For an example, when G is a (separable) lca group, the group (Fourier) transform $F_G: L^2(G) \to L^2(\Gamma)$ can be shown to intertwine two irreducible (unitary) representations of H on $L^2(G)$ and on $L^2(\Gamma)$; here H is the Heisenberg group associated with G. This essentially characterizes F_G and leads to a factorization of F_G into a product of three unitary operators. When specialized to the finite group Z/rs, where Z is the group of integers and r, s are integers > 1, the Cooley-Tukey FFT algorithm is obtained [28, 29]. This algorithm is also derived in Section III.4 below by a more direct group-theoretic procedure.

II.5 SAMPLING THEOREMS

Of the relative handful of significant mathematical theorems inspired by signal processing requirements, the most prominent is the so-called Sampling Theorem. Actually there is now a whole genre of results, loosely called sampling theorems, that pertain to the recovery of either deterministic or random signals from certain discrete information ('samples'). A rather complete survey of these results (up to 1977) has been given by Jerri [31]. Here we shall just look at a couple of prototypes in order to once again point out a group-theoretic setting.

For a deterministic signal the intuitive idea is that complete specification of the signal from a sequence of equally-spaced sample values is possible, provided that the signal does not vary too rapidly. This last phrase is taken to mean that the frequency spectrum of the signal is eventually zero. The relation between an upper bound on the spectrum and a sample spacing sufficient for recovery is a reciprocal one: if the spectrum vanishes outside the interval $[-\tau\ \tau$, then the signal is uniquely specified by its set of sample values taken at the points $\{k\alpha : k = 0, \pm 1, \pm 2, \ldots\}$, where $0 < \alpha \leqslant \pi/\tau$. In fact, we have the famous sampling formula

$$f(t) = \sum_{-\infty}^{\infty} f(k\pi/\tau)\ \mathrm{sinc}[\tau(t - k\pi/\tau)] \quad , \qquad \text{(II.25)}$$

valid for functions $f \in L^2(R)$ whose Fourier transform vanishes (ae) beyond the interval $[-\tau\ \tau]$; here

$$\mathrm{sinc}(x) = \begin{cases} 1 & , \quad x = 0 \\ \dfrac{\sin x}{x} & , \quad x \neq 0 \end{cases} \quad .$$

This formula is associated with the names of Cauchy (1841), Whittaker (1915), Nyquist (1928), Kotelnikov (1933), and Shannon (1949), who introduced and rediscovered it in various contexts. In addition to its well-known utility in communications theory (A/D, D/A conversion), formula (II.25) serves as the basis for a variety of numerical approximation procedures. In this setting the formula is known as the cardinal series expansion of f; see the survey by Stenger [32].

Proofs of the sampling formula are Fourier-analytical in nature. The nicest one proceeds from the Plancherel theorem, which establishes that the Fourier transform $f \rightarrow \hat{f}$ can be extended from $L^1(R) \cap L^2(R)$ to be a unitary operator on $L^2(R)$. (This theorem, of course, remains valid if R is replaced by an lca group.) The so-called Paley-Wiener space PWE_τ then is defined as the image of the (inverse) Fourier transform of the subspace of $L^2(R)$ consisting of those functions which vanish ae outside of $[-\tau, \tau]$. As is well known, the space PWE_τ is that subspace of $L^2(R)$ whose elements f can be extended to the complex plane so as to be entire functions of exponential growth there:

$$PWE_\tau = \{ f \epsilon L^2(R) : |f(z)| \leqslant c \exp(\tau |z|) \} \quad .$$

This Hilbert space of analytic functions, equipped with the reproducing kernel

$$K(z, w) = \frac{\tau}{\pi} \operatorname{sinc} \tau(z - \overline{w}) \quad , \tag{II.26}$$

is thus rich in structure and its elements serve as mathematical models of 'bandlimited signals.'

The convergence in Equation (II.25) results from the fact that the shifted sinc functions are the (inverse) Fourier transform of the standard expotential basis

$$\left\{\frac{1}{\sqrt{2\pi}} \exp\,(i\pi kt/\tau){:}k = 0, \pm 1, \pm 2, \ldots\right\}$$

on the interval $[-\tau\ \tau]$. Because this basis is orthonormal and the Fourier transform is unitary, the expansion (II.25) is just an expansion in an orthonormal basis in PWE_τ. However, by making use of the reproducing kernel properties of the space PWE_τ, one can also see that the convergence in Equation (II.25) is actually uniform on horizontal strips in the complex plane.

None of the foregoing analysis directly involves any group theory. Yet in view of the translations evident in formula (II.21), we might expect that a group theoretic version should exist. Certainly it is not particularly difficult to extend the sampling formula to several variables, that is, to produce a formula analogous to (II.25) that is valid for bandlimited functions on R^n, $n \geqslant 1$ [33]. [Actually, such generalizations date back to work of Parzen (1956) and Peterson and Middleton (1962).] And, in fact, as we shall indicate momentarily, a rather complete generalization of the formula exists for any lca group [34].

Before doing so, however, we briefly consider another type of extended sampling formula by asking whether it is possible to replace the sinc function in Equation (II.25) by other functions. That is, we are looking for expansions of the form

$$f(t) = \sum_{k=-\infty}^{\infty} f(\alpha k)\, \phi(t - \alpha k) \quad , \tag{II.27}$$

(again valid for $f \epsilon L^2(R)$, with the constraint that $\hat{f}$ have compact support. Convergence in Equation (II.27) should at least be in the metric of $L^2(R)$ and perhaps be uniform, if we are lucky. This kind of sampling formula might be more useful than the traditional one because of more favorable behavior of the function ϕ. For example, the smoother $\hat{\phi}$ is, the more rapidly ϕ will decay, and then fewer terms on the right side of Equation (II.27) will be needed to adequately approximate f(t).

The key condition for an expansion of this form to hold is that of reciprocal relationship between the size of the support of f and sample spacing α. Suppose that we can find open sets V, W satisfying

$$\operatorname{supp}(\hat{f}) \subset V \subset \overline{V} \subset W \quad ,$$

and that W is so small that the translates

$$\left\{ (w + 2\pi k/\alpha : k = 0, \pm 1, \pm 2, \ldots \right\}$$

are pairwise disjoint. Then there exists an infinitely differentiable function ψ such that $\psi = 1$ on V and $\psi = 0$ outside of W. One can now show the following [35]:

(a) the measure of W is finite;

(b) hence $\hat{f} \epsilon L^1 \cap L^2(R)$ and so f must be continuous;

(c) with ϕ the inverse Fourier transform of ψ, the series in (II.23) converges uniformly to f.

Returning now to the problem of extending the classical sampling formula (II.25) to a group context, we let G be an arbitrary lca group. The role of the sampling points $\left\{ k\pi/\tau \right\}$ will be played by a discrete subgroup H of G, and the goal, for a given $f \epsilon L^2(G)$, is an expansion of the form

$$f(g) = \sum_{h \epsilon H} f(h)\, \phi(g - h) \tag{II.28}$$

for a suitable function ϕ. As is to be expected, the matter depends on the size of the support of the group (Fourier) transform $\hat{f}$. We want it to be small relative to H, and we make this precise by introducing the annihilator A of H: $A = H^{\perp} = \left\{ \gamma \epsilon \Gamma : <h, \gamma> = 1, h \epsilon H \right\}$. As an example, when G = R and $H = \alpha Z$ for some fixed $\alpha > 0$, then $\Gamma =$ R and $A = (2\pi/\alpha)Z$ (here Z is the group of integers). Now we shall require that f vanish ae outside of an open set Ω which has the property that its translates $\left\{ \Omega + \gamma : \gamma \epsilon A \right\}$ are pairwise disjoint. With this setup one can now establish the following, originally proved in [34] under the further assumption that the subgroup A also be discrete:

(a) the (Harr) measure of Ω is a positive number $\beta^2 < \infty$;

(b) the restrictions of the character functions $\beta^{-1} <h, \cdot>$, $h \epsilon H$, to Ω form an orthonormal basis for $L^2(\Omega)$;

(c) hence, if the function ϕ is defined as the inverse Fourier transform of the characteristic function of Ω, then ϕ is continuous and positive definite, and the set of translates $\left\{ \beta^{-1}\phi(\cdot - h) : h \epsilon H \right\}$ is an orthonormal set in $L^2(g)$.

(d) f is continuous and the expansion (II-28) holds both uniformly on g and in the metric of $L^2(G)$.

Thus this argument basically parallels that already available for the classical case here G = R. But it reveals a little more; for instance, that a function on R may have an unbounded spectrum and yet still be completely determined by its values at the discrete sampling instants $\left\{ k\alpha \right\}$.

Finally, we want to indicate that analogous results are valid for random signals defined on groups. For stationary processes defined R, the corresponding sampling formulas go back to

Balakrishnan [36] and Lloyd [37]. In general, if $\{x_g:g\epsilon G\}$ is a weakly stationary process on an lca group G, as defined in Section II.2, and H is a closed subgroup of G, we say that $\{x_g\}$ is determined by its samples $\{x_h:h\epsilon H\}$ if the Hilbert space H_x is the closed linear span of $\{x_h\}$. Then one can show that if the support of the spectral measure of the process $\{x_g\}$ has the property that its translates by the members of the annihilator of H are pairwise disjoint, the process is determined by its samples along H in the above sense. This result can in turn be fairly easily extended to cover (strongly) harmonizable processes on G as defined in Section II.4 [35].

In the classical case ($G = R$, $H = \alpha Z$) much effort has been devoted to establishing sampling formulas similar to (II.25) or (II.27) that would converge, or at least be summable, in one sense or another to a given random process. For example, if the series in (II.25) is truncated to terms with $|k| \leqslant N$, if f denotes a weakly stationary process on R whose spectral measure is supported by the interval $[-\tau + \gamma, \tau - \gamma]$ for some $\gamma > 0$ (the 'guard band' assumption), then for $|t| \leqslant \pi/2\tau$ the mean square error $e_N(t)$ in approximating the random variable f(t) by the truncated series obeys an inequality of the form $e_N(t) \leqslant c(t)/N$, with $\lim c(t) = 0$ as $t \to 0$ [38].

Mean square convergence can in fact be established for sampling expansions of a large class of nonstationary processes on R which are bandlimited in a suitable sense. See for example [35] again, and its references, for a precise statement. A formula similar to (II.26) is obtained (with $H = \alpha Z$, as usual), and shown to converge almost surely, as well. For these results the key technical requirement of the process is that the (two-dimensional) Fourier transform of the covariance function of the process exist as a distribution in a certain Sobolev space on R^2, and that its support be contained in an open set whose translates by the points $\{k\alpha^{-1}(1, 1):k \epsilon Z\}$ are pairwise disjoint.

Returning once more to the general group context of the third paragraph above, we remark that it is possible to establish a sampling expansion for (strongly) harmonizable processes defined on G, given certain assumptions on the process and the subgroup H along which the process is sampled. Namely, it is supposed that H is an infinite closed discrete cyclic subgroup of G, and that the spectral measure of the process $\{x_g:g\epsilon G\}$ [as defined below Equation (II.17)] has its support contained in an open subset of $\Gamma \times \Gamma$ whose translates by the points (α, α), for α in the annihilator of H, are pairwise disjoint. The generalized sampling formula then appears as

$$x_g = \lim_{n\to\infty} \sum_{h\epsilon H} s_n(h)\, c_g(h) x_h, \quad g\epsilon G \quad ,$$

where the c_g are numerical coefficients, and the s_n are uniformly bounded functions of finite support on H that converge to the characteristic function of the identity element of H [35]. These arise from an extension to the group setting of a classical summation method for Fourier series. It would appear that some work remains to be done in this area before the exact conditions for a valid sampling formula are obtained, and the variety of relevant summability methods is specified.

REFERENCES

1. R. Holmes, "Mathematical Foundations of Signal Processing," SIAM Rev. **21**, 361-388 (1979).

2. W. Rudin, *Fourier Analysis on Groups* (Wiley, New York, 1962).

3. P. Masani, "Orthogonally Scattered Measures," Adv. Mathematics **2**, 61-117 (1968).

4. S. Dossou-Gbete, "Analyse Spectrale des Fonctions Aleatoires Generalisees Stationnaires du Second Order: Theoreme d'Isomorphisime de Kolmogorov," Cahiers du C.E.R.O. **22**, 159-173 (1980).

5. B. Brainerd and R. Edwards, "Linear Operators Which Commute with Translations. I: Representation Theorems, J. Austral. Math. Soc. **6**, 289-327 (1966).

6. P. Masani and M. Rosenberg, "When Is an Operator the Integral of a Given Spectral Measure?," J. Funct. Anal. **21**, 88-121 (1976).

7. R. Larsen, *An Introduction to the Theory of Multipliers* (Springer-Verlag, New York, 1971).

8. S. Cambanis ard E. Mazry, "On the Representation of Weakly Continuous Stochastic Processes," Inf. Sci. **3**, 277-290 (1971).

9. G. Muraz, "Operateurs Subordonnés a une Mesure Spectrale" (C.R.A.S. Paris 289, 1979), p. A271-273.

10. L. Koopmans, *The Spectral Analysis of Time Series* (Academic Press, New York, 1974).

11. M. Rao, "Harmonizable Processes: Structure Theory," L'Enseign. Math. **28**, 295-351 (1982).

12. S. Cambanis and B. Liu, "On Harmonizable Stochastic Processes," Inf. Control **17**, 183-202 (1970).

13. H. Hurd, "Testing for Harmonizability," IEEE **IT-19**, 316-320 (1973).

14. A. Miamee and H. Salehi, "Harmonizability, V-Boundedness and Stationary Dilation of Stochastic Processes," Indiana Univ. Math. J. **27**, 37-50 (1978).

15. P. Fillmore, *Notes on Operator Theory* (Van Nostrand Reinhold, New York, 1970).

16. P. Woodward, *Probabilty and Information Theory with Applications to Radar*, 2nd ed. (McGraw-Hill, New York, 1965).

17. A. Rihaczek, *Principles of High-Resolution Radar* (McGraw-Hill, New York, 1969).

18. S. Kay and F. Boudreaux-Bartels, "On the Optimality of the Wigner Distribution for Detection," Proc. ICASSP, 1985, pp. 1017-1020.

19. W. Schempp, "Radar Ambiguity Functions and the Linear Schrodinger Representation," in *Anniversary Volume on Approximation Theory and Functional Analysis*, P. Butzer *et al.*, eds. (Birkhauser-Verlag, Basel, 1984), pp. 481-491.

20. L. Auslander and R. Tolimieri, "Radar Ambiguity Functions and Group Theory," SIAM J.Math. Anal. **16**, 577-601 (1985).

21. A. Oppenheim, "Speech Spectrograms Using the Fast Fourier Transform," IEFE Spectrum 7, 57-62 (1970).

22. L. Cohen, "Generalized Phase-Space Distributions," J. Math. Phys. 7, 781-786 (1966).

23. F. Boudreaux-Bartels, "Time-Frequency Signal Processing Algorithms: Analysis and Synthesis Using Wigncr Distributions, Ph.D. Thesis, Rice University, 1984.

24. H. Weyl, *Theory of Groups and Quantum Mechanics* (Dover, New York, 1931).

25. P. Cartier, "Quantum Mechanical Commutation Relations and Theta Functions," in Proc. Symp. Pure Math., A. Borel and G. Mostow, eds. American Math. Soc., Providence, 1966, pp. 361-383.

26. *Special Functions: Group Theoretical Aspects and Applications*, R. Askey *et al.*, eds. (Reidel, Dordrecht, 1984).

27. W. Schempp, "Radar Ambiguity Functions, Nilpotent Harmonic Analysis, and Holomorphic Theta Series," in [26], pp. 211-260.

28. L. Auslander, "A Factorization Theorem for the Fourier Transform of a Separable Locally Compact Abelian Group," in [26], pp. 261-269.

29. ___________ and R. Tolimieri, "Is Computing with the Finite Fourier Transform Pure or Applied Mathematics?," Bull. Amer. Math. Soc. **1**, 847-897 (1979).

30. R. Howe, "On the Role of the Heisenberg Group in Harmonic Analysis," Bull. Amer. Math. Soc. **3**, 821-843 (1980).

31. A. Jerri, "The Shannon Sampling Theorem — Its Various Extensions and Applications," Proc. IEEE **65**, 1565-1596 (1977).

32. F. Stenger, "Numerical Methods Based on Whittaker Cardinal, or Sinc Functions," SIAM Rev. **23**, 165-224 (1981).

33. K. Chen and C. Yang, "On n-Dimensional Sampling Theorems," Appl. Math. Comp. 7, 247-257 (1980).

34. I. Kluvanek, "Sampling Theorem in Abstract Harmonic Analysis," Mat.-Fiz. Cas. **15**, 43-48 (1965).

35. A. Lee, "Sampling Theorems for Nonstationary Random Processes," Trans. Am. Math. Soc. **242**, 225-241 (1978).

36. A. Balakrishnan, "A Note on the Sampling Principle for Continuous Signals," IRE **IT-3**, 143-146 (1957).

37. S. Lloyd, "A Sampling Theorem for Stationary (Wide Sense) Stochastic Processes, " Trans. Am. Math Soc. **92**, 1-12 (1959).

38. F. Beutler, “On the Truncation Error of the Cardinal Sampling Expansion,” IEEE **IT-22**, 568-573 (1976).

39. J. Blum and B. Eisenberg, “A Note on Random Measures and Moving Averages on Non-Discrete Groups”, Ann. Prob. **1**, 336-337 (1973).

III. DATA PROCESSING OVER FINITE GROUPS

We now want to return to the context of Section II.1, which should be reread at this time, and work through in greater detail some of the mathematical aspects of linear data processing, as defined there. Hence, throughout this chapter we will model the observations as a random element in (equivalently, a probability measure on) a finite dimensional Hilbert space. Our emphasis will be on the choice of 'good' orthonormal coordinate systems to facilitate the data processing task at hand. Any such choice leads immediately to a corresponding unitary operator by which to transform the data. Particular attention will be paid, as already promised in Section II.1, to those unitary operators which are group (Fourier) transforms wrt some underlying group structure. The interesting trade-offs here concern the choice of group, whose structure then determines the transform complexity and hence computational efficiency, the nature of the signal and noise statistics, the estimation errors or distortion, and the amount of data compression.

III.1 UNITARY OPERATORS FOR DATA PROCESSING

Following the preceding introduction and the background of Section II.1 we now consider data in the form of an element y belonging to a finite dimensional Hilbert space $\underline{Y}$. The precise nature of y and $\underline{Y}$ is not too important initially, but typically it will be the case that y is a column vector $(y_1, \dots, y_N)^t$ so that $\underline{Y}$ is the space C^N of all complex N-tuples with the usual algebraic operations and inner product

$$<u, v> \sum_{j=1}^{N} u_j \overline{v}_j \quad .$$

Alternatively, y might model an image and so would appear after preprocessing as a matrix $[y_{ij}]$; then Y would be the space of all such matrices of the same dimension with the weak (Hilbert-Schmidt) norm derived from the inner product

$$<u, v> = \mathrm{tr}(uv^*) \quad .$$

Of course, such data could be rearranged ('stacked') to form a column vector.

Now before proceeding to the analysis we have to consider what we might want to do with, or learn from, this data. From the several possible generic goals listed in Section I.1 we will consider here only three:

- Dimensionality reduction
- Transform coding
- Signal estimation.

The first of these, often termed *feature selection*, consists of replacing y by its projection on a subspace of Y. We thus represent the data by fewer parameters, often with the intent of submitting this reduced data to a pattern classifier. The basic issue is then to choose the optimal subspace, having fixed its dimension and an error criterion. The latter will depend on the prior information available concerning the data generating mechanism; for instance, this may be a known, or estimated, probability distribution.

Now any projection on $\underline{Y}$ of rank d is unitarily equivalent in many ways to the projection of C^N onto C^d which simply drops the last N-d components. So, as we apply various unitary transforms U to the data, $U:\underline{Y} \rightarrow C^N$, the resulting first d components constitute the possible d-dimensional data reductions. Although these will generally be suboptimal wrt any given error criterion, some of them may be more efficiently computed than the optimal transform. This will be the case if U has a fast algorithm, that is, one where the computational effort in obtaining Uy is less than the expected $O(N^2)$ floating point operations ('flops'). Such a U is generically called a *fast unitary transform* ('FUT'); the FFT is the most famous example. The point here is that if the error made in choosing the leading d components of Uy, where U is a FUT, does not greatly exceed the minimum possible error, the computational savings may offset the slightly higher error. Then, larger size data blocks at a higher sampling rate could be processed, resulting in an increase in overall system performance.

Use of Fourier or other unitary transforms as preprocessing for pattern classification dates back to the 1969-71 time period; the sources [48, 49] may be consulted along with their many references.

A similar situation occurs in *transform coding*, one of the principal methods of data compression [1, 2]. This is a collection of techniques aimed at reducing the amount of signal space necessary for a given signal, where the components of signal space are, generally, physical space, time, and bandwidth. Data compression is widely employed in the fields of speech coding, telemetry, television, facsimile, and data base access. In general, the problem is the efficient transmission of the information contained in a multidimensional source signal, and the idea is to eliminate the redundancy in the signal prior to encoding.

A schematic of transform coding is given in Figure III-1. The choice of the first transform is driven by the requirements of preserving information while returning uncorrelated components. This prewhitened data may then be individually quantized and encoded. After transmission through the channel and subsequent decoding, a final transform is applied to restore, as far as possible, the original signal.

The major goal in the transform coding technique of data compression is to be able to employ fewer bits in quantization than would be the case if transforms were not first applied, or if the data were treated separately rather than in blocks. If the number of available bits is fixed then they should be allocated so as to minimize some measure of overall distortion. In fact, the selection and efficient quantization of the transformed components for storage or transmission is at least as important in terms of overall system performances as the choice of blocksize and transform. However, in this report the latter is of primary interest. Let us just note here that generally an optimal choice of bits per source symbol, given a certain channel capacity, depends

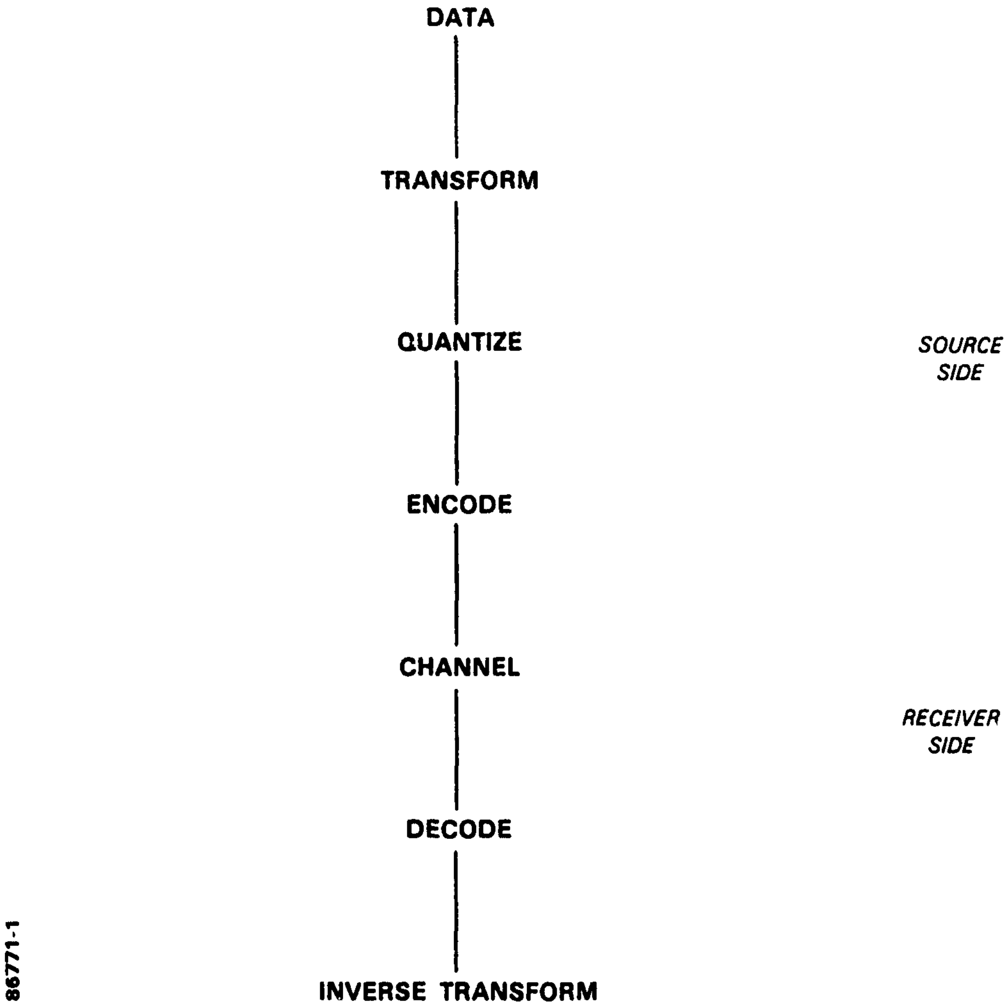

Figure III-1. Transform coding schematic.

on the variance of the components of the transformed data vector. If this variance fails to exceed a threshold the component is dropped (that is, 0 bits are assigned); otherwise the number of bits assigned depends on the variance and a distortion function.

The topic of signal estimation was already introducted in Section II.1, where it was concluded that for zero-mean signal and noise the optimal linear data processor is the Wiener filter of Equation (II.3). In greater detail, conserving the notation $y = s + \eta$ of Equation (II.1), we note that if an estimate

$$\hat{s} = L \cdot y \tag{III.1}$$

of the signal s is formed by an arbitrary linear operator L acting on the data y, then the mean squared error e_L is given by

$$\begin{aligned} e_L &= E(\|s - \hat{s}\|^2) \\ &= \mathrm{tr}(C_{s-\hat{s}}) \\ &= \mathrm{tr}(C_s - C_{sy}L^* - LC^*_{sy} + LC_yL^*) \\ &= \mathrm{tr}(LC_yL^*) - 2 \text{ re } \mathrm{tr}(LC^*_{sy}) + \mathrm{tr}(C_s) \\ &= <LC_y, L> - 2 \text{ re} <L, C_{sy}> + <C_s, I> \quad , \end{aligned}$$

where $<A, B> = \mathrm{tr}(AB^*>$ is the Hilbert-Schmidt inner product on the space of operators on Y. In this form the error e_L can readily be minimized wrt L; the optimal operator L is the Wiener operator W as defined in Equations (II.3). Further, the minimum error e_W is given by

$$\begin{aligned} e_W &= \mathrm{tr}\ (WC_yW^* + C_s) - 2 \text{ re tr } (WC^*_{sy}) \\ &= \mathrm{tr}\ (C_{sy}C_y^{-1}C^*_{sy} + C_s) - 2 \text{ re tr } (C_{sy}C_yC_y^{-1}C^*_{sy}) \\ &= \mathrm{tr}\ [C_s - C_s(C_s + C_\eta)^{-1}C_s] \quad , \end{aligned} \tag{III.2}$$

where we have used $C_{sy} = C_s$ and $C_y = C_s + C_\eta$ for uncorrelated signal s and noise η. We note that these arguments could be generalized to an infinite dimensional setting provided that η is interpreted as a second-order weak random variable

Since the signal-to-noise estimation problem is so fundamental, we make a brief excursion at this point. Consider the case where s = A(x), that is, we have a linear inverse problem as discussed following Equation (II.4). If x belongs to a Hilbert space X so that A:X→Y, and L:Y→X is any potential solution operator, then the mean squared error e_L (averaged over the noise distribution) can be expressed as

$$e_L = \|x - LAx\|^2 + \mathrm{tr}(LC_\eta L^*) \quad , \tag{III.3}$$

a formula which remains valid in the infinite dimensional case provided that L is restricted to be a Hilbert-Schmidt operator. Now, the point is that if no prior information is available concerning x, there is no way to choose a single operator L to make e_L uniformly small for every $x \epsilon X$. In order to obtain a unique solution, therefore, some additional constraint must be imposed. A classical restriction is to make the estimate $\hat{x} = L(y)$ unbiased, so that $LA = I_X$. Then it results that

$$L_0 = (A^*C_\eta^{-1} A)^{-1} A^* C_\eta^{-1} \tag{III.4}$$

gives the minimum mean squared error. This operator is sometimes known as the Gauss-Markov estimator. Technical requirements for its existence are that C_η be invertible (obviously!), and that A have a trivial nullspace.

Now, as research on James-Stein and ridge estimators has shown [3, Chapter II], a willingness to accept some bias can yield a smaller mean square error. Suppose then that we drop the unbiased restriction and instead consider, as we have been doing before, a prior distribution on x. Then the optimal (Wiener) estimator has the form

$$W = C_x A^* \, (A C_x A^* + C_\eta)^{-1}$$

$$= (C_x^{-1} + A^* C_\eta^{-1} A)^{-1} \, A^* C_\eta^{-1} \quad . \qquad \text{(III.5)}$$

The first formula above remains valid in the infinite dimensional case provided that the covariance operator C_x of the prior is nuclear (or, trace class); otherwise, we restrict to the finite dimensional case and assume, for the second formula, that this covariance is positive definite, hence invertible. We can note from this second formula that as prior knowledge of x becomes more diffuse in the sense that $\|C_x\| \to \infty$, the Wiener operator W converges to the Gauss-Markov operator L_0 defined by Equation (III.4).

These remarks aside, let us now return to the earlier case (dim $Y<\infty$, $A = I$) with Wiener filter $W = C_s(C_s + C_\eta)^{-1}$ and error e_W given by Equation (III.2). We note that in general this operator has no particular structure, except in the important but special case where the noise η is white, so that C_η is a scalar matrix. In that case W is a positive (semi-) definite operator. Another sufficient condition for W to be hermitian is that C_s commute with C_η. Failing this, we fall back on the theorem that the product of hermitian operators is hermitian if and only if it is normal. From this we conclude that W is normal <===> W is hermitian <===> $C_\eta C_s^{-1}$ is hermitian, the last provided that the signal covariance is nonsingular. In any event, it is not particularly easy to compute with W, and this difficulty leads to the concept of generalized Wiener filtering.

As in the preceding cases of data reduction and coding we consider a preliminary transformation of the data y by a unitary operator $U{:}\underline{Y} \to C^N$. We then multiply this transformed data by a matrix A and inverse transform. That is, in the notation of Equation (III.1), our estimates have the form

$$\hat{s} = L \cdot y = U^* A U \cdot y \quad . \qquad \text{(III.6)}$$

Any such transformation, for $U \neq I$, is called a generalized Wiener filter [4]. What is interesting here is that if we work through the minimization of the error $e_L = E(\|s - \hat{s}\|^2)$ again, we find that, for fixed U, the optimal choice of A is UWU*, where W is the original Wiener filter, and that the minimum value of e_L is e_W, as given in Equation (III.2). Thus the minimum error turns out to be independent of the choice of transform U. In particular, we are free to choose U so that the optimal A has a simple form.

Since we must make some error no matter what we do, the real issue is how to coordinate the choice of U with some suboptimal but simple form of the matrix A. For example, a natural first question is whether, for some U, the associated optimal A is diagonal. Since A = UWU*, an equivalent question is whether the Wiener filter W is normal. As was observed earlier it certainly is, provided that the noise covariance is scalar or, more generally, commutes with the signal covariance. When W is normal by virtue of the noise η being white, any unitary operator that diagonalizes it is at the same time one diagonalizing the signal covariance C_s; these are called (discrete) Karhounen-Loeve transforms (DKLTs) of the signal.

Let us now summarize the pros and cons of the use of the DKLT for signal estimation; unfortunately, there are more of the latter. First, it need not always exist — we have to assume commutativity of noise and signal covariance operators. Secondly, it changes with any change in the signal statistics. Third, there is no reason to expect it to be a FUT, in general, so that as the data blocksize increases we have an increasingly lengthy eigenvector computation to accomplish. For these reasons the role of the DKLT in data processing is more that of a benchmark rather than a viable numerical procedure.

This being said, we can now suggest the main concepts of suboptional Wiener filtering. We. consider the generalized Wiener filters of Equation (III.6), with U defined independently of the signal statistics and A chosen to be 'simple'. This last term is deliberately a little vague; we have in mind that A should be diagonal or at least close to diagonal terms that are nonzero. Further, U should be an FUT so that the computational effort is reduced. The essential trade-off, then, is between filter complexity and error, for different statistical signal environments.

With this extended motivation for the use of unitary transforms, particularly FUTs, for several generic data processing purposes behind us, let us think a little about such operators from a general point of view: what they do, or ought to do, to be useful, and how they can be constructed. First of all, we recall that unitary transforms are, in effect, changes of basis. That is, if $U:\underline{Y}\rightarrow C^N$ is unitary and we write

$$U(x) = \begin{bmatrix} \alpha_1 \\ \cdot \\ \cdot \\ \cdot \\ \alpha_N \end{bmatrix} \quad ,$$

then each α_i depends linearly on x, so that there is a vector $u_i \epsilon Y$ with

$$\alpha_i = <x_i, u_i> \quad .$$

Since, by assumption,

$$||x||^2 = ||U(x)||^2 = \sum |\alpha_i|^2 \quad , \tag{III.7}$$

it easily follows that $\{u_1, \ldots, u_N\}$ is an orthonormal basis (for short, a *frame*) in Y. So, the effect of U is to pick out the coordinates of an element of Y wrt a particular frame. We will call these coordinates the *spectral coordinates* of x wrt $u_1, \ldots, u_N$.

When x is a data vector obeying some zero-mean probability law on Y, its spectral coordinates wrt the frame $\{u_i\}$ are random variables on Y. In this case the numbers

$$\gamma_i = E(|< x, u_i >|^2) \quad , \quad i = 1, \ldots, N \tag{III.8}$$

constitute the *power spectrum* of x wrt the given frame. Although the relative size of the γ_i will vary from one frame to another, we have the identity

$$\sum_{i=1}^{N} \gamma_i = E(||x||^2) \quad , \tag{III.9}$$

independently of frame. Note that the numbers γ_i can also be interpreted as the values $< C_x u_i, u_i >$, where C_x is the covariance operatoı of the random variable x. If, in particular, the frame $\{u_i\}$ is chosen so as to diagonalize C_x, then γ_i is just the spectrum of this operator, and the number defined in Equation (III.9) is seen to be $tr(C_x)$. This particular set γ_i is sometimes called the *normal* power spectrum, and a corresponding frame u_i of eigenvectors is a Karhounen-Loeve basis for Y (relative to the law of x).

At this point we know that unitary transforms preserve the energy of observed data vectors [this is the import of Equation (III.7), and also the total signal power [the value in Equation (III.9)]. This latter number is sometimes referred to also as the (statistical) bandwidth of the data. We may also note, without proof, that unitary transforms preserve the relative, or cross-entropy between a pair of random vectors in Y. That is, if x, y are random vectors in Y with distributions p, q, respectively, whose cross-entropy H(p, q) as defined in Equation (I.6) is finite, then

$$H(x, y) \equiv H(p, q) = H(Ax, Ay)$$

where A is any unitary or, more generally, nonsingular linear operator on Y. So, in this precise sense, unitary transforms preserve the information about one random vector contained in another. This sort of result goes back to the early work of Shannon, Kolmogorov, Gelfand and Yaglom in information theory [5]; see also the more recent and more detailed work of Rosenblatt-Roth [6].

We are still left with the task of selecting a unitary transform to fit a particular data processing task, or, equivalently, a suitable set of spectral coordinates. We have seen that we can't distinguish between such transforms on the basis of power or information-theoretic criteria. What we can expect, however, is to differentiate on the basis of the statistical behavior of the individual spectral coordinates. Specifically, desirability of unitary transforms increases with their ability to decorrelate these coordinates and to pack most of the signal power into a small number of them. This latter phrase means roughly that for M much less than N the sum $\gamma_1 + \ldots + \gamma_M$ should be near to the total signal power of Equation (III.9).

This last goal is an example of a data processing task, the performance of which can be measured by a function F of the power spectrum coordinates, namely,

$$F(\gamma_1, \ldots, \gamma_N) = \sum_{j=1}^{M} \gamma_j \quad .$$

The goal is achieved by a unitary operator U for which F is maximized. Now it is fairly direct to show that if $\{u_j\}$ is a Karhounen-Loeve basis for Y and $\{v_i\}$ is any other frame, then its power spectrum $\underline{\lambda} = (\lambda_1, \ldots, \lambda_N)$ is related to the normal power spectrum $\underline{\gamma} = (\gamma_1, \ldots, \gamma_N)$ associated with u_j by

$$\underline{\lambda} = A\underline{\gamma}$$

where A is the orthostochastic matrix $[|<u_i, v_j>|^2]$. By Birkhoff's theorem, A is a convex combination of permutation matrices:

$$A = \sum \alpha_k P_k$$

and so

$$\underline{\lambda} = \sum \alpha_k (P_k \underline{\gamma}) \quad .$$

That is, all possible power spectra of a given second order probability measure lie in the convex hull of the permutations of the normal power spectrum. The import of this observation, which goes back to J. Pearl [7], is that the merit function F is maximized by a DKLT of the data. Of course, this nice theoretical result ignores computational realities.

A second example of such merit functions occurs when the processing goal is efficient transform coding; for example, the spectral coordinates are to be transmitted through a binary channel of some fixed capacity C bits/symbol. In terms of a given distortion function ϕ, decreasing and convex, and an assignment of B_j bits to the jth coordinate (the latter being assumed independent here), the merit function F becomes

$$\min \left\{ \sum \phi (B_j) \gamma_j : B_j \geqslant 0, \sum B_j = C \right\} \quad .$$

From this it follows that F is, in fact, a certain linear function of $(\gamma_1, \ldots, \gamma_N)$ [8].

Another kind of merit function, not of the above form, is needed when trying to rank unitary transforms by their ability to decorrelate the spectral coordinates of a probability measure or, perhaps, a 'small' class of such. If, as usual, we deal with a covariance matrix C_x and a unitary transform U, the spectral coordinates obey a probability law whose covariance matrix is $\Lambda = UC_xU^*$. Its diagonal entries γ_{ii} are the components of the power spectrum. Hence an appropriate figure-of-merit would be some measure of the magnitude of the off-diagonal entries γ_{ij}; for instance

$$F(\Lambda) = \frac{1}{N} \sum_{i \neq j} |\gamma_{ij}|^2 \quad . \tag{III.10}$$

This last data processing task, namely, to choose a unitary transform to approximate a DKLT of a particular covariance (or class thereof), has led, over the past decade, to some interesting work [9, 10] on asymptotic properties of various spectral representations, which we summarize briefly here. The general idea, as in so much of statistical theory, is to study the behavior of certain approximations to 'truth' as the sample size (blocklength, here) becomes infinite.

Specifically, suppose given a sequence $\{U_N : N = 1, 2, \ldots\}$ of N-dimensional unitary transforms, and, for each N, a family $\mathbf{C}_N$ of positive-semidefinite N-dimensional matrices. Each class $\mathbf{C}_N$ is intended to consist of possible covariances of observed data. Also, let F_N be a non-negative function defined on the space of all N × N matrices such that $F_N(\Lambda) = 0$ if Λ is diagonal. F_N is intended to measure how far a matrix is from being diagonal; it could be defined, for instance, by Equation (III.10). We can then agree to say that the transform sequence U_N asymptotically decorrelates the family $\mathbf{C}_N$ if, for each sequence C_N , where $C_N \epsilon \mathbf{C}_N$,

$$\lim_{N \to \infty} F_N(U_N C_N U_N^*) = 0 \qquad \text{(III.11)}$$

Thus, if F_N is defined by Equation (III.10) (hereafter referred to as the 'standard case'), then a sufficient condition for (III.11) to hold is that each off-diagonal entry γ_{ij} of $U_N C_N U^*_N$ satisfy $|\gamma_{ij}|^2 = o(1/N)$, as $N \to \infty$.

The underlying motivation for the foregoing abstract set-up is the desire to rapidly process large blocks of data by first applying an FUT to the data vector and then treating the resulting spectral coordinates independently for coding/compression purposes. Since the exact data statistics are rarely known, we must expect the processing to be effective over a class $\mathbf{C}_N$ of possible covariance matrices. The prototypical example is that where the data is obtained as a segment of a discrete weakly stationary process, so that $\mathbf{C}_N$ consists of Toeplitz matrices, and $\{U_N\}$ is the sequence of N-dimensional DFTs. Then it is known [11] that if the process is restricted to have square-summable covariance sequence, the DFT sequence will asymptotically decorrelate all the corresponding (Toeplitz) covariance matrices.

In general, if $C \epsilon \mathbf{C}_N$ and $U_N C U_N^* = D$ (a diagonal matrix) in the sense that $F_N(U_N C U_N^*)$ is small, and we set $C' = U^*_N D U_N$, then C' is diagonal wrt the frame associated with U_N and $C' \approx C$, at least in the standard case, by unitary equivalence. So, asymptotical decorrelation of the family $\{\mathbf{C}_N\}$ by the transform sequence $\{U_N\}$ is equivalent to even better approximation of the matrices in the U_N frame. When U_N is the N-dimensional DFT, the corresponding class of diagonable matrices consists of circulants.

So far we have not mentioned the rate of convergence in Equation (III.11). At least in the standard case this rate is interesting for two reasons. First, for a given family $\{\mathbf{C}_N\}$ of covariances, convergence rates allow us to compare the performance of different transform sequences. This might permit us to decide, for instance, whether certain data might best be processed with Fourier, Walsh, Haar, cosine, or yet other transforms. Second, by assigning numerical performance criteria for various data processing tasks, we might be able to bound or estimate the performance degradation resulting from the use of these various fast but suboptimal approximations to the DKLT.

As an illustration, suppose that $\mathbf{C}_N$ is the class of covariance matrices corresponding to first order Markov processes. That is, $C \epsilon \mathbf{C}_N$ means

$$C = [\rho^{|i-j|}], \qquad \text{(III.12)}$$

for $1 \leqslant i,j \leqslant N$ and $0 < \rho < 1$; ρ originates as the correlation coefficient between adjacent samples. Here the spectrum and Karhounen-Loeve basis can be explicitly obtained [4]. Further, it can be shown that, not only does the DFT sequence asymptotically decorrelate these covariances, but so does the popular discrete cosine transform (DCT) [13], and in fact it is 'better' than the DFT in this context. That is, while the rate of convergence in Equation (III.11) is the same for both transform sequences, the DCT error is strictly less than the DFT error, for every choice of ρ [12]. A related remark here is that the DCT is now known to be derived from the limiting form (as $\rho \to 1$) of the Karhounen-Loeve basis for C in Equation (III.12) [14].

At present, the DFT, the DCT, and several other kinds of unitary transforms have been embedded into some general theories. We have in mind here on the one hand the Gauss-Jacobi transforms of Yemini and Pearl [10], which are based on the classical convergence of Gaussian quadratures derived from orthogonal polynomials, and, on the other hand, the sinusoidal transforms of Jain [15]. These latter are the eigenvector frames of a parameterized family of Jacobi-like matrices of the form

$$J(k_1, k_2, k_3) = \begin{bmatrix} 1 - k_1\alpha & -\alpha & - & k_3\alpha \\ -\alpha & 1 & - & \\ \vdots & & & \\ & - & 1 & -\alpha \\ k_3\alpha & - & -\alpha & 1 - k_2\alpha \end{bmatrix}$$

These sources should be consulted for pertinent details. We are now going to turn our attention in a different direction, to a consideration of unitary transforms arising from a group-theoretic setting.

III.2 GROUP ALGEBRAS AND REPRESENTATIONS

Having attempted to motivate the use of unitary operators in various data processing algorithms, we now specialize to the case where these operators are the group transforms of various finite groups. In this section we will quickly review the relevant group theory background and then look at the structure of group transforms in the next section.

Recall that we are dealing with data presented as random element of an N-dimensional Hilbert space $\underline{Y}$. In the previous section we pointed out the connection between unitary operators on $\underline{Y}$ and frames in $\underline{Y}$; in particular, a unitary map from $\underline{Y}$ onto C^N gives the (spectral) coordinate vector associated with a particular frame B in $\underline{Y}$. This association effectively realizes $\underline{Y}$ as the discrete sequence space $\ell^2(B)$. Now the key idea underlying the rest of this report is the frame B may have some additional structure — for instance, it might be a group.

Here are two simple examples when N = 4. First of all, there are only two distinct (nonisomorphic) groups of order 4: the cyclic group C_4 and the Klein 4-group D_2. Both are abelian; indeed, any group of order p or p^2, p a prime, is necessarily abelian. Assume that $\underline{Y} = C^4$ and consider the frame $B_1 = \{u_1, u_2, u_3, u_4\}$, where

$$u_1 = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}, \quad u_2 = \begin{bmatrix} 1 \\ w \\ w^2 \\ w^3 \end{bmatrix}, \quad u_3 = \begin{bmatrix} 1 \\ w^2 \\ w^4 \\ w^6 \end{bmatrix}, \quad u_4 = \begin{bmatrix} 1 \\ w^3 \\ w^6 \\ w^9 \end{bmatrix}$$

and $w = \exp(2\pi i/N) = i$, here. Under the operation of componentwise multliplication, this frame is easily seen to be a group isomorphic to C_4. Similarly, the frame $B_2 =\ v_1, v_2, v_3, v_4$ defined by

$$v_1 = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}, \quad v_2 = \begin{bmatrix} 1 \\ -1 \\ 1 \\ -1 \end{bmatrix}, \quad v_3 = \begin{bmatrix} 1 \\ 1 \\ -1 \\ -1 \end{bmatrix}, \quad v_4 = \begin{bmatrix} 1 \\ -1 \\ -1 \\ 1 \end{bmatrix}$$

is again a group under componentwise multiplication, this time isomorphic to D_2.

Now, while these 'group-frames' may seem more or less natural, we can, in fact, realize any group of order N as a frame in $\underline{Y}$. Namely, let G be such a group and let its Haar measure m_G be normalized so that $m_G(G) = 1$; that is, $m_G(g) = 1/N$ for each $g \epsilon G$. Then the space $L^2(G)$ is N-dimensional and hence congruent with $\underline{Y}$. It contains the frame $\{\sqrt{N}e_g : g \epsilon G\}$ where e_g is the indicator function of $\{g\}$. The image of this frame in $\underline{Y}$ under any congruence T is again a frame, and it is claimed that this frame can be given a group structure under which it is isomorphic to G. This claim easily follows from the facts that the space $L^2(G)$ is an algebra under convolution as multiplication, and that the product $e_g e_h = e_{g*h}$, for g, $h \epsilon G$. Then the group structure on the frame $\{\sqrt{N}e_g\}$ is carried over to its image frame in $\underline{Y}$ so that T becomes an isomorphism.

The upshot of these observations is that it is possible to model the data as a random element of rather special Hilbert spaces of the form $L^2(G)$, where the order of G is the blocklength of the data. As was noted back in Section 1.2, while Hilbert spaces of the same dimension are abstractly equivalent, they individually possess widely different realizations as sets of functions or operators. In the present case spaces $L^2(G)$ have a very rich structure, going far beyond that permitted by the usual Hilbert space axioms. This structure can be revealed by several different approaches: group representations, harmonic analysis, and Banach algebras, to name three. These theories are, of course, very powerful and extensive, and servc to similarly describe the structure of $L^2(G)$ for general compact topological groups G, and many others additionally. In this section we will just review those structural aspects that seem relevant to data processing applications.

First, as earlier noted, $L^2(G)$ is an algebra under convolution multiplication:

$$f_1 * f_2 (g) = \frac{1}{N} \sum_{h \epsilon G} f_1 (gh^{-1}) f_2 (h) \quad .$$

Thus, since G is finite, $L^2(G)$ is set-theoretically identical with the so-called 'group algebra' of G. When G is not of finite order, that term is more commonly applied to the space $L^1(G)$. There is also an involution $f \rightarrow f^*$ defined by

$$f^*(g) = \overline{f(g^{-1})} \quad .$$

These operations are tied together with the inner product

$$<f_1, f_2> \int_G f_1 f_2 \, dm_g \equiv \frac{1}{N} \sum f_1(g) \overline{f_2(g)}$$

by the formula

$$<f_1 * f_2, f_3> = <f_2, f_1^* * f_3> \quad . \qquad \text{III.13}$$

Thus $L^2(G)$ is at the same time a Banach algebra and a Hilbert space. Such spaces, with property (III.13), are called H*-algebras; their structure has been described by Ambrose [16] and recounted by Loomis [17]. The basic fact is that such a space is uniquely expressible as an orthogonal direct sum of its minimal (closed) two-sided ideals, each of which is isomorphic to a full matrix algebra.

Since we are ultimately interested in using the group (Fourier) transform, it is more natural here to introduce (unitary) representations of G as the key technical tool for the study of $L^2(G)$. The representation theory of finite groups, due originally to Frobenius, and developed by Schur, Burnside, Weyl, and many others, is purely algebraic and is described in detail in many sources; for instance, the books of Keown [18] and Serre [19]. It has also been generalized, largely intact, to the case of compact groups, where analytic techniques predominate and harmonic analysis (generalized Fourier series) is often the focus. This material, originating with the Peter-Weyl theorem in 1927 (briefly outlined in Section 1.2) is available in, for example, the books of Edwards [20] and Naimark-Stern [21], and in the Hewitt-Ross treatise [22].

We now review (rapidly) just those aspects of representation theory necessary to reveal the proper setting for the group (Fourier) transform ana the structure of the group algebra. In general, a *representation* of a group G is a strongly continuous homomorphism T from G into the group of invertible operators on some complex topological vector space V. Since our major interest is in the case where G is finite, the continuity of T is trivial and without essential loss we may take $\dim(V)<\infty$. If (. , .) is an arbitrarily assigned inner product on V, the formula

$$<u, v> = \int_G (T(g)u, T(g)v)\, dm_G(g)$$

defines a new inner product on V in which the T(g) are unitary operators. So we may restrict attention to unitary representations of G.

A representatior. T is *irreducible* if there is no nontrivial subspace of V that is invariant under all the operators T(g), $g\epsilon G$. An easy induction shows that every representation is completely reducible, in the sense of being a direct sum of irreducible representations. Thus these latter are the building blocks of the general theory, although finding a complete list of them for a given group may be very difficult.

Some operator criteria for irreducibility of unitary representations are the following: since a subspace of V is invariant under $\{T(g): g\epsilon G\}$ exactly when its orthogonal projection commutes with each T(g), it follows that T is irreducible if and only if the commutant of $\{T(g)\}$ consists only of scalar operators (Schur's lemma). Similarly, one can show that the algebra span $\{T(g):g\epsilon G\}$ is the space L(V) of all operators on V exactly when T is irreducible (Burnside's theorem).

Every irreducible representation of G is finite dimensional. (Proofs of this fundamental fact for general compact groups are often based on the eigenstructure of a compact hermitian operator, but it can be made to follow only from Schur's lemma [23].) Of these, the simplest examples occur when dim V = 1. In this case we effectively are looking at homomorphisms of G into the circle group T and we have already referred to such mappings as characters. When G is abelian it turns out that the characters form a group Γ under natural operations and that G is canonically isomorphic to the dual of Γ; this is an instance of the Pontryagin duality theorem which in fact remains valid for general lca groups. In this abelian case it is further true that all irreducible representations are one-dimensional, hence characters of G. Finally, it can be verified that the characters constitute a frame in the group algebra $L^2(G)$; hence G and Γ are of the same order. The group Γ is called the *dual group* of G and one of the tasks of the nonabelian theory is to find a suitable substitute for it that continues to shed light on the structure of G and its group algebra.

Two representations $T:G\rightarrow L(V)$ and $S:G\rightarrow L(W)$ are *equivalent* if there is an isomorphism $A:V\rightarrow W$ such that $A\cdot T(g) = S(g)\cdot A$, $g\epsilon G$. By this notion, inessentially different representations are collected together in equivalence classes. If, as we assume, V and W are finite dimensional Hilbert spaces and S, T are equivalent unitary representations, then S and T are actually unitarily equivalent; the proof utilizes the polar decomposition of the isomorphism A.

One way to distinguish between equivalence classes of (unitary) representations is by an extended notion of character. Namely, if $T:G \to L(V)$ is any representation of G, define its character χ_T by

$$\chi_T(g) = \mathrm{tr}[T(g^{-1})] \quad . \tag{III.14}$$

One easily checks that

$$\chi_T(e) = \dim(V),\ \chi_T(g^{-1}) = \overline{x_T(g)} \quad , \tag{III.15}$$

for T unitary, and that characters of equivalent representations coincide. Conversely, by decomposing a given representation into a direct sum of irreducible representations, one can show that two representations with the same character are equivalent. Thus, a representation is 'characterized', up to equivalence, by its characteı.

By calculations based on Schur's lemma one shows that the characters associated with inequivalent irreducible representations form an orthonormal set in $L^2(G)$. Actually, the norm of the character associated with any representation is always at least one, and equals one exactly in the irreducible case.

Now as a replacement for the dual group, we define the *unitary dual object* Γ to be the set of equivalence classes of irreducible unitary representations of G. From the foregoing remarks we could equally well take Γ to be the set of associated characters — the 'irreducible characters.' By the orthonormality property this set, denoted $\{\chi_1, \ldots, \chi_r\}$, is part of a frame in $L^2(G)$, so that $r \leqslant \mathrm{ord}(G)$. In fact, more is known: if we denote by d_i the dimension of the space of any representation associated with χ_i, then we have the *Burnside formula*

$$\mathrm{ord}(G) = d_1^2 + \ldots + d_r^2 \quad . \tag{III.16}$$

Also, each integer d_i divides the order of G and also the index of the center of G in G. This last remark is usetul provided that the center of G is nontrivial; this is the case for instance, if G is a p-group, that is, ord(G) is a power of some prime p. The center then contains at least p elements.

Moving back now to the group algebra $L^2(G)$, its center consists of the so-called class functions f defined by the condition

$$f(hgh^{-1}) = f(g), \qquad g, h \in G.$$

These are just the functions on G that are constant on each of the conjugacy classes of G. Examples are the characters of any representation of G. It turns out that the irreducible characters of G span, and hence constitute a frame for, this space of class functions. Hence the cardinality r of G is also the number of conjugacy classes of G. The rxr matrix whose (i, j) entry is the value of the ith irreducible character on the jth class of G is often called the *character table* of G.

The orthonormality of the irreducible characters in $L^2(G)$ serves another purpose. It was earlier noted that any representation T of G is a direct sum of irreducibles. Using the characters of all these representations we can write an explicit formula for this, namely

$$T = \sum_{i=1}^{r} < \chi_T, \chi_{T_i} > T_i \quad , \tag{III.17}$$

where each T_i is a representative from an equivalence class in Γ.

Before moving on to the structure of the group algebra we offer a few comments to draw together some loose ends. We have noted that the characters of one-dimensional unitary represention coincide with the characters introduced previously as homomorphisms of G into the circle group. These latter may be called *group characters* and we know they form an abelian group Γ. When G is abelian Γ determines G by duality. Otherwise, there are complications. The commutator subgroup G′ of G is nontrivial. The abelian quotient group G/G′ has the same dual group as G. Hence the group characters only help us understand the commutative structure of G. When G is nonabelian there is at least one irreducible (unitary) representation of G on a space of dimension >1. Further, there may be no nontrivial group characters at all; this would be the case, for instance, if G has no nontrivial normal subgroups (G is then said to be *simple*). In any event, when G is not abelian the dual object, also denoted Γ above, does not have a natural group structure and does not generally determine G [22, p. 57]. A successful substitute for Γ was introduced by Tannaka in 1939 for general compact groups and later axiomatized by Krein (1949) and Kelley (1963). This is the space T_G spanned by the coordinate or representative functions on G, that is, functions of the form $f(g) = <T(g^{-1})u, v>$, as T runs through the classes in Γ (also called trigonometric polynomials). A certain set of linear functionals on T_G turns out to admit a group structure under which it is compact and naturally isomorphic with G [22, Sec. 30].

As a brief aside we remark that a very active topic of research in the period 1959-1974 was the development of a general duality theory for noncompact and nonabelian locally compact groups. Contributions to this area were made by W. Stinespring, P. Eymard, J. Ernest, K. Saito, N. Tatsuuma, M. Takesaki, C. Akemann, and M. Walter, in rough chronological order. All this work makes substantial use of the theory of operator algebras and related functional analysis, and attempts to characterize a given locally compact group G in terms of a related space of functions on G or operators on $L^2(G)$. Thus the thrust here is in a different direction from much of the earlier work on locally compact groups which was concerned with decomposition of specific unitary representations, generally of infinite dimension, of such a group.

The structure of the group algebra $L^2(G)$ is a consequence of properties of the set $\chi_1, \ldots, \chi_r$ of irreducible characters of G. As elements of $L^2(G)$ these characters are hermitian $(\chi_i{}^* = \chi_i)$ and obey the orthogonality relations

$$\chi_i * \chi_j = \delta_{ij} \, d_i^{-1} \, \chi_i \quad .$$

Hence the (two-sided) principal ideals J_i generated by the χ_i are orthogonal subspaces of $L^2(G)$; for this, the formula

$$<f, h> = f^{**} h(e)$$

is helpful. The set $\{J_1, \ldots, J_r\}$ of these ideals is exactly the set of minimal two-sided ideals in $L^2(G)$ and we have the decomposition

$$L^2(G) = J_1 + \ldots + J_r \tag{III.18}$$

with corresponding orthogonal projection $P_i : L^2(G) \rightarrow J_i$ given by

$$P_i(f) = f * d_i\ \chi_i \quad . \tag{III.19}$$

This ideal can also be defined as consisting of all functions on G of the form

$$f_A(g) = \mathrm{tr}[A \cdot T_i(g)^{-1})] \quad , \tag{III.20}$$

where A is an arbitrary operator on the space V_i of the irreducible representation T_i. If $e_n^{(i)}$ is a frame in V_i then a frame for the ideal J_i is $\left\{f_{mn}^{(i)}\right\}$, where

$$f^{(i)}_{mn}(g) = \sqrt{d_i} < T_i(g^{-1}) \cdot e^{(i)}_m, e^{(i)}_n >. \tag{III.21}$$

The correspondence $f_A \leftarrow A$ defined by Equation (III.20) above sets up an isomorphism between J_i and the full operator algebra $L(V_i)$, $i = 1, \ldots, r$. An essentially inverse isomorphism may be achieved by expanding $f \epsilon J_i$ in the frame elements $f_{mn}^{(i)}$,

$$f = \sum c^{(i)}_{mn} f^{(i)}_{mn} \quad ,$$

and making the $d_i x d_i$ matrix $[c^{(i)}_{mn}]$ correspond to f. Under the first correspondence, the central character χ_i in J_i maps into the identity of $L(V_i)$, hence χ_i serves as the identity of J_i.

All these results generalize more or less directly to the case where G is a separable compact group. Expansion of an arbitrary $f \epsilon L^2(G)$ in all the frame bases (III.21) as T_i runs through a (countable) complete set of irreducible (finite dimensional) unitary representations of G yields the group Fourier series for f. Such an expansion reduces to the classical Fourier expansion of a periodic function when G is the circle group.

We'll close this section with a comment about the regular representation(s) of a group G. Its decomposition into irreducible components contains the essence of harmonic analysis on groups, and provides much motivation for the foregoing results. The *right regular representation* of G on $L^2(G)$ is defined by

$$R(g_o)\ f(g) = f(gg_o)$$

There is also a left regular representation which is unitarily equivalent to R. These unitary representations are natural extension of the translation group $\left\{\tau_g : g \epsilon G\right\}$ discussed in Section II.3 for lca groups G.

A subspace M of $L^2(G)$ is *invariant* if $R(g)M \subset M$, $g \epsilon G$, and a general goal of harmonic analysis is to decompose $L^2(G)$ into a direct sum of such subspaces. In this context, if we fix an index i, $1 \leqslant i \leqslant r$, and consider functions f_A as defined by Equation (III.20), we see that

$$R(g)\ f_A = f_{T_i(g)A} \quad , \quad g \epsilon G \quad ,$$

so that the ideals J_i in the decomposition (III.19) are invariant.

In another approach to harmonic analysis one can begin with the representation R and attempt to decompose it into a sum of irreducible representations of finite dimension. We know this is possible for groups of finite order, and in fact, it continues to be true for all compact groups. But in general, a locally compact group need not have any nontrivial finite dimensional unitary representations, irreducible or not (certain connected semisimple Lie groups, for example). Actually, in our elementary situation it follows from Equation (III.17) that every irreducible representation occurs in the regular representation with a multiplicity equal to its dimension. This too continues valid when G is compact. In the general locally compact case, whether or not a particular irreducible unitary representation occurs in the regular representation depends on further assumptions about the group. If, for example, there is a Plancherel measure on Γ, (as there is when G is unimodular and type I), then points in its support are exactly those contained in the regular representation.

When G is finite, as we are assuming in this chapter, there is a natural isomorphism between the group algebra and the commutant of the regular representation. This is the map that assigns to each $f \epsilon L^2(G)$ the operator of convolution with f. Now the communtant of any finite dimensional representation is a direct sum of full matrix algebras (Schur's lemma again), so in this fashion it is possible to derive anew the basic decomposition (III.18) of the group algebra.

III.3 GROUP TRANSFORMS

The importance of Fourier methods in signal processing was briefly recalled and emphasized in Section I.2, and the Fourier transforms were defined in Equation (I.1) and (I.2). At an abstract level, which we will not stress, one can think of the Fourier transform together with an accompanying Plancherel theorem as an explicit solution to the general problem of decomposing the regular representation into irreducible components. The search for such a formula for various noncompact and nonabelian groups has been a major theme of group representation theory. However, reviewing this is not germane to the present discussion. We will continue to look at the case of finite groups, where all such existence questions are trivial, and to view the Fourier transforms as simply a particular kind of unitary transform with interesting properties and structure, and possible relevance to discrete data processing.

We can ease into the definition of the Fourier transform through the idea of extending a given group representation to the group algebra. Let G be a compact group and $T:G \rightarrow L(V)$ a finite dimensional representation. For $f \epsilon L^1(G)$ we define

$$T(f) = \int_G f(g)\ T(g)\ dm_G(g) \qquad . \tag{III.22}$$

This extends T to be a continuous representation of the algebra $L^1(G)$ by operators on V:

$$\begin{aligned} &\|T(f)\| \leqslant B\|f\|_1 \quad , \\ &T(f*h) = T(f)\cdot T(h) \quad , \end{aligned} \tag{III.23}$$

if B is a bound on $\left\{\|T(g)\|:g\epsilon G\right\}$. When V is a Hilbert space and the representation is unitary, then

$$T(f^*) = T(f)^* \quad ,$$

and the extended T is a *-representation of the algebra $L^1(G)$. Since G is compact and of Haar measure one, $L^2(G) \subset L^1(G)$ via $||f||_2 \leqslant ||f||_1$, so that T is also a norm-decreasing *-representation of $L^2(G)$. When G is finite, of order N, Equation (III.22) simply defines T by linearity:

$$T(f) = \frac{1}{N} \sum_{g \epsilon G} f(g)\ T(g) \quad . \tag{III.24}$$

Three brief comments about this construction are appropriate. First, it is reversible, so that we in fact have a one to one correspondence between unitary representations of G and nontrivial *-representations of $L^1(G)$. Second, the extended T has the same commutant as the group representation; hence if one is irreducible, so is the other. Third, using the invariance of Haar measure, one easily verifies that

$$T[R(g_o)f] = T(f)T(g_o)^* \quad , \tag{III.25}$$

where R is the (right) regular representation of G. Hence from this formula and the earlier Equation (III.23) we see that these *-representations of $L^2(G)$ send convolution products and tranlations into certain operator products. These are, of course, generalizations of familiar valuable properties of Fourier transforms, which we next define.

From now on we restrict attention to finite groups G, denoting ord(G) by N and the unitary dual object of G by Γ. We let $T_i{:}G \rightarrow L(V_i)$ be a representative of the ith class of Γ, with $\dim(V_i) = d_i$, $i \leqslant i \leqslant r \leqslant N$. Recall the r = N only for abelian groups. Each T_i extends to $L^2(G)$ by formula (III.24) and we let T be the product map

$$T = (T_1, \ldots, T_r) : L^2(G) \rightarrow \prod_{i=1}^{r} L(v_i) \quad . \tag{III.26}$$

It is a consequence of the structure theory for $L^2(G)$ recounted in the previous section that each T_i defines, by restriction, an isomorphism between J_i in Equation (III.18) and $L(V_i)$. Indeed, T_i is surjective by Burnside's theorem, and injective by the formula

$$f^*\chi_i = \mathrm{tr}\ [T_i(f) \cdot T_i(\cdot)^*] \quad ; \tag{III.27}$$

recall from Equation (III.19) that the left side above is $d_i^{-1}f$ if $f \epsilon J_i$. This last fact implies that

$$T_i(\chi_i) = d_i^{-1}\ I_i,$$

where I_i is the identity operator on V_i. Now we can see that the inverse of T_i on $L(V_i)$ is defined by

$$T_i^{-1}\ (A) = d_i\ f_A \quad , \tag{III.28}$$

where f_A was defined by Equation (III.20). It suffices to check this for $A = T_i(g_o)$, $g_o \epsilon G$:

$$
\begin{aligned}
T_i(d_i f_A) &= \frac{d_i}{N} \sum_g tr[T_i(g_o) \cdot T_i(g^{-1})]\ T_i(g) \\
&= \frac{d_i}{N} \sum_g tr\ [T_i(g_o g^{-1})]\ T_i(g) \\
&= \frac{d_i}{N} \sum_h tr\ [T_i(h^{-1})]\ T_i(hg_o) \\
&= di\ (\frac{1}{N} \sum_h tr[T_i(h^{-1})]\ T_i(h)\ T_i(g_o) \\
&= d_i\ [T_i(\chi_i)]\ T_i(g_o) \\
&= I_i\ T_i(g_o) = T_i(g_o) \quad .
\end{aligned}
$$

From Equation (III.28) in turn we see that

$$T^{-1}\ (A_1, \ldots, A_r) = f \quad ,$$

where

$$P_i(f) = d_i\ f_{A_i}, \qquad 1 \leqslant i \leqslant r \quad .$$

This yields a complete analysis and synthesis of an arbitary function $f \epsilon L^2(G)$:

$$f(g) = \sum_{i=1}^{r} d_i\ tr[T_i(g^{-1}) \cdot T_i(f)] \tag{III.29}$$

The *group (Fourier) transform* is the mapping T defined by Equation (III.24) and (III.26), with *inverse transform* defined by Equation (III.29) above. We will hereafter refer to T as simply the group transform, denoted $F_G(f)$, sometimes $\hat{f} = F_G(f)$, to emphasize its dependence on the group G.

We conclude the first half of this section with several comments about this definition, along with one more key property (the Plancherel theorem); then we'll look at some examples and discuss the complexity issue.

First, because we are limited to finite groups, there are no convergence or integrability issues and the inversion formula (III.29) is always valid. Second, the definitions of convolution, representation and extended *-representation, character of a representation, and choice of Haar measure on the group must all be carefully and consistently chosen to make the various important properties of the group transform work out, especially those describing the transform of convolutions (III.23) and translations (III.25), and inversion (III.29). Other definitions appear in the literature: our f*h may be another author's h*f or (1/N) f*h, our character another's conjugate character, etc. All of these are equally valid, as long as they are consistently followed. We also might recognize that there is a certain nonuniqueness in our definition of the group transform, in that it depends on a specific choice of representation T_i from each class in Γ. However, this nonuniqueness is really inessential as both the dimensions d_i and characters χ_i are well defined, and hence so are the projections P_i defined by Equation (III.19), etc.

At this point we should recall that our major theme of this chapter is the use of unitary transforms for data processing. We now want to see that the group transform as just defined is indeed unitary. This involves checking that its range has a Hilbert space structure, and that

$$\|f\|_2 = \|\hat{f}\|, \qquad f \epsilon L^2(G) \quad . \tag{III.30}$$

We do this by noting that the range of F_G is, according to the definition of T in (III.26), just the direct product of the operator algebras $L(V_i)$, $i = 1, \ldots, r$. Since each V_i is finite dimensional, each of these algebras is actually an H*-algebra under the Hilbert-Schmidt inner product [$<A, B> = tr(AB^*)$]. Hence the product space is also an H*-algebra under the inner product

$$<(A_i, \ldots, A_r), (B_i, \ldots, B_r)> \quad = \quad \sum_{i=1}^{r} <A_i, B_i> \quad .$$

We will denote this product space $L^2(\Gamma)$, since it can be also thought of as the space of all functions on Γ whose value at the ith class is an operator on V_i. If a measure ρ on the discrete space Γ is defined by assigning the value d_i to the ith class, then $\phi = (A_1, \ldots, A_r) \epsilon L^2(\Gamma)$ has the norm

$$\|\phi\|^2 \quad = \int_{\Gamma} \|(A_1, \ldots, A_r)\|^2 \, d\rho$$

$$= \sum_{i=1}^{r} d_i \ tr\,(A_i A_i^*) \quad . \tag{III.31}$$

What we claim is that $F_G: L^2(G) \rightarrow L^2(\Gamma)$ is unitary in that the relation (III.30) holds when $\|\hat{f}\|$ is defined by (III.31). Specifically, we have the *Plancherel theorem* for the finite group G:

$$<f, h> = \sum_{i=1}^{r} d_i \ tr\,[T_i(f)\ T_i(h)^*] \tag{III.32}$$

for all f, $h \epsilon L^2(G)$. (By contrast, the formula obtained from Equation (III.29) by setting g = e (group identity) is often called Plancherel's formula. It relates $f \epsilon L^2(G)$ to its scalar (not operator!)-valued Fourier transform, thus assigning to f the scalar function on G whose value at the ith class is $<f, \chi_i>$, $1 \leqslant i \leqslant r$. Of course, this transform is one-to-one only when G is abelian.)

In essence, this formula is just a reflection of the orthogonal decomposition (III.18). If, for example, we make use of the frames $\{f^{(i)}_{mn}\}$ in J_i defined by Equation (III.21), then

$$\|f\|_2^2 \quad = \sum_{i=1}^{r} \ \sum_{m,n=1}^{d_i} |<f, f^{(i)}_{mn}>|^2$$

$$= \sum_{i=1}^{r} d_i \, [\sum_{m,n} |<T_i(f)\, e^{(i)}_n, e^{(i)}_m>|^2]$$

$$= \sum_{i=1}^{r} d_i \operatorname{tr} [T_i(f)\ T_i(f)^*] \quad ,$$

by definition of trace. By polarization, this is equivalent to Equation (III.32). Alternatively, one can simply expand the right hand side of Equation (III.32). and employ the formula

$$d_i\ \chi_1 + \ldots + d_r\ \chi_r = N \cdot \mathrm{id} \quad , \tag{III.33}$$

where 'id' is the identity element of the group algebra, that is, the indicator function of the group identity. This latter formula arises again from the decomposition (III.18) and the projection formula (III.19) since, for any $f \epsilon L^2(G)$,

$$f = \sum_{i=1}^{r} P_i(f) = \sum_{i=1}^{r} f^* d_i\ \chi_i$$

$$= f * \ (\sum_{i=1}^{r} d_i\ \chi_i) \quad .$$

It also arises directly from the inversion formula (III.29) by the substitution $f = \mathrm{id}$.

This completes the background development in Fourier analysis for finite groups. We note that all these formulas generalize rather directly to general compact groups. There, as in Equation (III.31), a Plancherel measure exists and assigns finite mass d_i to each class $\{T_i\}$ in Γ. The Plancherel theorem for lca groups and for certain locally compact nonabelian groups is discussed in References [11] and [13], respectively, for Chapter I. This theorem, and the accompanying inversion formula all specify quantitatively the way in which the irreducible representations of a group G permit a harmonic analysis of square-integrable functions on G.

Before moving on to the complexity of the group transform and its data processing applications, we digress briefly to the topic of positive definite functions. We will just consider these on finite groups; although much (but not all) of what we say extends to general locally compact groups. By definition, a complex-valued function ϕ on the group G is *positive-definite* if for all subsets $g_1, \ldots, g_n$ of G, the matrix

$$[\phi(g_i\ g_j^{-1}]$$

is positive semidefinite. In particular, it follows that

$$| \phi(g) \leqslant \phi(e) \geqslant 0 \quad ,$$

$$\phi(g^{-1}) = \overline{\phi(g)} \quad ,$$

for all $g \epsilon G$. If $U{:}G \rightarrow L(V)$ is a unitary representation of G on a Hilbert space V, and $v \epsilon V$, then

$$\phi(g = < U(g)v,\ v >$$

defines a positive-definite function of G and, in fact, all such functions arise in this way.

Positive-definite functions are of interest for data processing because they occur as autocorrelation functions in one of two ways. Suppose first that f_1, f_2 are functions on G. Their *cross-correlation function* is defined by

$$\rho_{1,2}(x) = \int \overline{f_1(g)}\, f_2(gx)\, dm_G(g)$$

$$= \frac{1}{N} \sum_{g \epsilon G} \overline{f_1(g)}\, f_2(gx) \quad ,$$

for $x \epsilon G$. It is easy to check that

$$\rho_{1,2}(x) = f_1^* * f_2(x) \quad , \quad x \epsilon G$$

and hence that in the case $f_1 = f_2$,

$$\rho_{1,1} = f_1^* * f_1, \tag{III.34}$$

the *autocorrelation function* of f_1. As elements of the group algebra, functions of this latter form are called *hermitian squares*, and are positive-definite by virtue of

$$\rho_{1,1}(x) = <R(x)\, f_1, f_1> \quad ,$$

where R is the right regular representation of G.

If the functions f_1, f_2 are thought of as sample functions of a stochastic process on G, the autocorrelation and cross-correlation functions are essential components of Wiener's generalized harmonic analysis, although this term is usually applied when G is the group of real numbers (and then the definition of $\rho_{1,2}$ must be modified to account for the infinite Haar measure of G).

For the second example we proceed as in Section II.2 and consider a stochastic process $\{x_g : g \epsilon G\}$ which is weakly stationary in that there is a unitary representation U of G on $L_0^2(P)$ with $x_g = U(g) \cdot x_e$, $g \epsilon G$. The function

$$\rho(g) = E(x_e\ x_e)$$

then satisfies

$$E(x_g\, \overline{x_h}) = \rho(h^{-1}g) \quad , \tag{III.35}$$

and may again be called an autocorrelation function. However, to distinguish between these two cases we will refer to it as the *covariance function* of the process. Clearly

$$\rho(g) = <x_g, x_e> = <U(g)\, x_e, x_e> \quad , \tag{III.36}$$

so that ρ is a positive-definite function on G. This key property of ρ is easily established because of the right choice of definition and the nontrivial characterization of positive definite functions mentioned above Finiteness of G is not at all needed for these results.

One other important example of positive-definite functions is the character χ_T, as defined by Equation (III.14), of a finite dimensional unitary representation T of G. The proof follows from the above characterization, again, and the fact that the positive-definite functions form a convex cone in the group algebra. This cone, denoted PD(G), is also closed under conjugation, involution, and products (that the product of positive-definite functions is again positive-definite is a nice application of Schur's result that the Hadamard product of positive semidefinite matrices is again positive semidefinite).

To see some other characterizations of positive-definite functions, let $S(G) = \{$ hermitian squares $\} = \{f{:}f = h^{*}{*}h,\ h \epsilon L^2\,(G)\}$, and $M_\ell(G)$, [resp., $M_r(G)$] $= \{f \epsilon L^2(G)$:left (resp., right) multiplication by f is a positive semidefinite operator$\}$. Then it is not difficult to show that

$$PD(G) = {}^{\prime}S(G) = M_\ell(G) = M_r(G),$$

where ${}^{\prime}S(G)$ means the dual cone of S(G), that is, ${}^{\prime}S = \{f{:}<s, f> \geqslant 0,\ s \epsilon S\}$.

Now for finite groups it can be shown that actually

$${}^{\prime}S(G) = S(G)$$

in other words, that S(G) is a self-dual cone in the group algebra $L^2(G)$. This is a consequence of the structure theory for $L^2(G)$ reviewed in Section III.2 together with the analogous fact for the H*-algebra L(V), V a finite dimensional Hilbert space [24].

Finally, we define a *positive* function to be one with a positive Fourier transform. That is, we set

$$P(G) = \{ f \epsilon L^2(G){:}\ \hat{f} \geqslant 0 \} \quad ;$$

this means that each operator $T_i(f)$ is a positive semidefinite on the ith-representation space when T_i runs through Γ. To complete our circle of characterizations of PD(G), we claim that

$$P(G) = M_\ell(G) \quad .$$

This can be seen in various ways. For example, the Plancherel theorem (III.32) implies that

$$<f^{*}h,\ h> = \sum_{i=1}^{r} d_i <T_i(f)\ T_i(h),\ T_i(h)> \quad ,$$

for each f, $h \epsilon L^2(G)$. Keeping in mind that the fact mentioned above that the positive semidefinite operators on a finite dimensional Hilbert space form a self-dual cone, it follows from $f \epsilon P(G)$ that $<f^{*}h,\ h> \geqslant 0$ for all $h \epsilon L^2(G)$ and so $f \epsilon M_\ell(G)$. Conversely, it is clear from basic properties of the group transform that $S(G) \subset P(G)$, and we already know, for finite groups G, that $M_\ell(G) = S(G)$.

The relation P(G) = S(G) can be considered as an analogue of the classical Fejer-Riesz theorem about non-negative trigonometric polynomials, this being essentially equivalent to the corresponding relation P(Z) = S(Z) for the integer group Z. However, not much more generality is possible; for example, $P(G) \neq S(G)$ when $G = Z \oplus Z$ [24].

For signal processing applications the most important of the above characterizations is PD(G) = P(G), showing that with each positive-definite function, in particular, with each autocorrelation and covariance function, is associated 'something positive'. In the more familiar case where G is abelian, this 'something' is just a function on the dual group Γ with non-negative real values. Guided by the more general case of lca groups it is more useful to think of this function as a measure on Γ. This brings us into conformity with the viewpoint of Section II.2 where, via Bochner's theorem for lca groups, the covariance function was viewed as the Fourier transform of a positive measure on the dual group. That measure was termed the spectral measure of the underlying stochastic process. It is the measure or its derivative, the spectral density function, that is the object of estimation procedures in the field of spectrum estimation. This is not an area that we are going to discuss except to note that prior to the more modern high resolution methods, a sample function was used to make estimates. Either the autocorrelation function was first estimated and then Fourier transformed to yield a spectrum estimate (Blackman-Tukey approach), or else a DFT was applied directly to the data and then a multiple of its squared magnitude served as the estimate (periodogram approach). There are many issues here that must be resolved before convergence of such estimates can be guaranteed, and techniques of time-domain windowing or frequency-domain smoothing play a key role [25, 26].

The dual relation between an autocorrelation or covariance function and its group transform is called the Wiener-Khinchine relation, as already noted following Equation (II.2). When the underlying group is not abelian the resulting transform is a (positive semidefinite) operator-valued function on the dual object Γ. Much of the rest of this report deals with tentative data processing applications of this general setup.

We first note a purely mathematical formula involving positive definite functions, and then give it a statistical interpretation. Let $\phi \in PD(G)$. Then ϕ defines a positive linear functional Φ on $L^2(G)$ by virtue of $PD(G) = {}^1S(G)$. That is, we have

$$\Phi(f) = \langle f, \phi \rangle$$

and

$$\begin{aligned} 0 \leqslant \Phi(f^* * f) &= \langle f^* * f, \phi \rangle \\ &= \langle f, f * \phi \rangle = \langle f, f \cdot R \rangle \end{aligned} \tag{III.37}$$

where the operator R on the space of the Fourier transforms $\hat{f}$ [as given in Equation (III.26)] has components $d_i T_i(\phi)$ in each $L(V_i)$, $1 \leqslant i \leqslant r$. When G is abelian, each $d_i = 1$ and this formula reduces to the familiar statement that convolution with a positive definite function transforms into the operator on $L^2(\Gamma)$ of multiplication by a positive function.

Now let a data vector be given, and considered as a random element in $L^2(G)$. We ask: when does the group transform F_G decorrelate this data? In other words, we ask: when is the covariance operator of F_G(data) a diagonal operator? Experience with classical transforms already warns us that this is a rather restrictive condition. For example, as is well known, the ordinary DFT (to be 'officially' defined in the next section) decorrelates a data vector if and only if the covariance matrix is a circulant. That result follows from the expression of a circulant as a polynomial in the shift (mod N) operator on $L^2(C^N)$.

In the general case, let the data be a realization of the weakly stationary process $\{x_g : g\epsilon G\}$ with covariance function ρ as defined in Equation (III.35), and $E(x_g) = 0$, $g\epsilon G$. Then the covariance operator on $L^2(G)$ is (essentially) the operator of right convolution with ρ. That is, for $f\epsilon L^2(G)$ interpreted as a linear functional of the data,

$$\begin{aligned} \mathrm{var}(f) &= E(|<x_g, f>|^2) \\ &= N^2 <f, f*\rho> \quad . \end{aligned} \tag{III.38}$$

Here, N = ord(G) as usual, and its appearance is due to the basic definition of convolution and choice of Haar measure on G. Also note, notationally, that the inner products in Equation (III.38) refer to $L^2(G)$ and not to a space of random variables, as in Equation (II.1) or (III.36). Equation (III.38) gives the statistical interpretation of Equation (III.37), and from the latter we also see the form of the diagonal operator R which serves as the covariance operator of the transformed data $\{x_g : g\epsilon G\}$. Namely, each component of R in the decomposition of Equation (III.26) has the form $d_i T_i(\rho)$, and each of these is a positive semidefinite operator on $L(V_i)$, $i = 1, \ldots, r$. There is therefore an eigenvector frame $e_j^{(i)} : j = 1, \ldots, d_i$ in each V_i for the operators $T_i(\rho)$ and hence the operator $R : A \rightarrow A\hat{\rho}$ is diagonal wrt the frame $\{e_j^{(i)} \otimes e_k^{(i)} : j, k = 1, \ldots, d_i, i = 1, \ldots, r\}$.

Once again, in the more familiar abelian case, there is a particularly nice version of Equation (III.38). Namely, in terms of the spectral measure μ on Γ,

$$\mathrm{var}(f) = \int_\Gamma |\hat{f}|^2 d\mu \quad .$$

What has just been shown is that the group transform F_G decorrelates any weakly stationary data $\{x_g : g\epsilon G\}$ in the sense that the covariance operator of $\{F_G(x_g)\}$ is a diagonal operator. This argument can be run backwards: if the transformed data has a diagonal covariance, then the original covariance C_x is right convolution with a positive-definite function ρ by our general theory. That is

$$<C_x a, b> = <a*\rho, b> \quad ,$$

for a, $b\epsilon L^2(G)$. Expanding both sides, we have

$$\begin{aligned} <C_x a, b> &= E(\overline{<x_g, a>} <x_x, b>) \\ &= \sum_g \sum_h a_g \overline{b_h} E(\overline{x_g} x_h) \quad , \end{aligned}$$

and

$$<a*\rho, b> = \frac{1}{N^2} \sum_g \sum_h \rho(g^{-1}h) a_g \overline{b_h} \quad ,$$

showing that

$$E(x_g \bar{x}_h) = \frac{1}{N^2} \rho(h^{-1} g) \quad ,$$

and hence that $\{x_g : g \epsilon G\}$ is weakly stationary.

Another way to phrase this conclusion is that the covariance operator of the data should be a *G-circulant*, that is, diagonal in the frame-basis for $L^2(G)$ defined by the group transform. When G is abelian this amounts to saying that this operator should be diagonal in the character basis for $L^2(G)$. This specializes to the classical case when G is cyclic and the group transform is the ordinary DFT; the latter will decorrelate a random vector if and only if the covariance matrix is a circulant.

III.4 TRANSFORM COMPLEXITY

Let's summarize where we stand in this survey of the use of finite groups for discrete data processing. In Section III.1 we discussed the general rationale of taking unitary transforms of a data vecto:. Then we noted that such a vector could be realized as an element of the group algebra oι groups whose order coincided with the data blocklength. This permitted us to bring to bear the general structure theory of discrete group algebras and, in particular, to define the associated group transform as a unitary operator. It is this type of operator that will be our focus for the remainder of this report; however, in keeping with its general level we will continue, for the most part, to avoid great detail in specific examples. That is more properly deferred to a more narrow and specialized study.

In earlier sections we have occasionally made reference to the notion of a 'fast' unitary transform (FUT) without attempting a definition. The general subject of fast transforms has been vigorously developed since 1965, when the Cooley-Tukey FFT algorithm appeared [27]; a fairly current view of the state of the art is given in the book [28]. Here we will just recognize two rather general approaches to the problem, which is, in essence, simply the efficient computation of the matrix-vector product Ux, where x may be of rather large dimension (e.g., several hundred or thousand). One is by a somewhat *ad hoc* collection of rules for manipulating the rows and columns of a unitary matrix, so as to preserve its unitary nature, and for building new larger unitary matrices from sets of smaller ones by recursive application of the Kronecker product operation. This methodology is described by Fino and Algazi [47]. The second approach is through the use of group theory (naturally restricted to group transforms!) and is sketched out next.

So, our problem is the computational complexity of the group transform operation

$$\hat{f} = F_G(f) \quad ,$$

where $f \epsilon L^2(G)$, and G is some group of order N. Let's see what's involved if we just proceed from the definitions by brute force. We'll first consider the relatively more simple case where G is abelian and then have a quick look at the general case.

Recall that when G is abelian the dual group Γ, also of order N, consists of N independent characters $\chi_1, \ldots, \chi_N$, which form a frame in $L^2(G)$. In this case the group transform is the unitary map from $L^2(G)$ to $L^2(\Gamma)$ defined, according to Equation (III.24), by

$$\hat{f}(\chi_k) = \frac{1}{N} \sum_{g \epsilon G} f(g)\, \chi_k(g) \quad , \quad k = 1, \ldots, N \quad . \tag{III.39}$$

If we employ as a conventional measure of computational complexity the number of (complex) multiplications and additions, we see that the complexity of directly computing $\hat{f}$ from Equation (III.39) is N^2 multiplications and $N(N - 1)$ additions (approximately, one of the 'χ's is identically one, and we ignore the 1/N factor). So we may say that F_G is a 'fast transform' if there is a numerical procedure for carrying out the computations in Equation (III.39) that requires 'significantly' fewer than $O(N^2)$ multiplications and additions. This definition is necessarily a little imprecise: some algorithms may be faster than others.

Now the key point is that whether or not a given group has a fast transform and, if so, how fast it is [relative to the $O(N^2)$ benchmark], depends on the subgroup structure of the group. Suppose that G is both abelian and decomposable, that is,

$$G = G_1 \times G_2 \tag{III.40}$$

for a pair of subgroups G_1 and G_2, of orders r and t, respectively. Then st = N and

$$\Gamma \equiv \Gamma_1 \times \Gamma_2 \tag{III.41}$$

where '≃' means 'isomorphic to'. This isomorphism is accomplished by $\chi \leftrightarrow (\phi, \psi)$, $\chi(g) = \chi[(g_1, g_2)] = \phi(g_1) \cdot \psi(g_2)$, for $\phi \epsilon \Gamma_1$, $\psi \epsilon \Gamma_2$. Hence we can rewrite the sum in Equation (III.39) as

$$\begin{aligned} N\,\hat{f}(\chi) &= \sum_{g \epsilon G} f(g)\,\chi(g) \\ &= \sum_{g_2 \epsilon G_2} \sum_{g_1 \epsilon G_1} f[(g_1, G_2)]\,\phi(g_1)\,\psi(g_2) \quad , \end{aligned}$$

and observe that the right hand side can be evaluated with a total of s - 1 + t - 1 complex multiplications and additions.

If one of the subgroups G_1, G_2 is itself decomposable, then the above process can be repeated. We conclude that if the abelian group G of order N factors as

$$G = G_1 \times \ldots \times G_n$$

with $\mathrm{ord}(G_i) = N_i$, then $F_G(f)$ can be computed in about

$$N \cdot \sum_{i=1}^{n} (N_i - 1)$$

complex multiplications and additions. In the important special case when all the groups G_i are isomorphic and of order p, the operations count above reduces to

$$(p - 1)\,N \log_p N \quad .$$

This value is particularly familiar when p = 2, and motivates the practical interest in working with data defined over groups whose order is a power of 2. In this context note that the optimization problem

$$\min \left\{ x_1 + \ldots + x_n : x_1 \ldots x_n = N,\ x_1 \geqslant 0 \right\}$$

is solved when all the $x_i = \sqrt[n]{N}$.

Before looking at some examples and cases where the group does not factor, we should first note that quite analogous reasoning establishes a similar reduction in complexity for the group transform over certain nonabelian groups. The basic assumption, once again, is the decomposability of the given group. Thus, as before, assume that G is of order N and that Equation (III.40) holds, with G_1 and G_2 of orders N_1 and N_2. Let $\{T_1, \ldots, T_r\}$ be a choice of irreducible representations from the classes of Γ. A fast algorithm for F_G depends on the proper generalization of Equation (III.41). This is achieved through use of the tensor product notion via the correspondence $T \leftrightarrow R \otimes S$, where R, S are irreducible representations of G_1, G_2, on Hilbert spaces U, V, respectively. This means that if $g = (g_1, g_2) \epsilon G$, T(g) is that operator on $U \otimes V$ whose value at $u \otimes v$ is $R(g_1)u \otimes S(g_2)v$.

Now the 'brute force' complexity of F_G can be obtained from the definition (III.24) with $T = T_1, \ldots, T_r$, successively, and Burnside's formula (III.16). We find a total of $Nd_i^2 - d_i + 1$ multiplications and $(N - 1)d_i^2$ additions for each i, $1 \leqslant i \leqslant r$, and therefore a total of $N^2 - \Sigma d_i + r$ multiplications, and $N(N - 1)$ additions; hence, just as in the abelian case, $O(N^2)$ operations in all. These counts are obtained by treating each $T_i(g)$ as a $d_i \times d_i$ matrix.

To get a fast algorithm for F_G we fix i, $1 \leqslant i \leqslant r$, and write

$$N\, T_i(f) = \sum_{g_2 \epsilon G_2} \sum_{g_1 \epsilon G_1} \left[f(g_1, g_2)\, R(g_1)\right] \otimes S(g_2) \quad , \qquad \text{(III.42)}$$

in accord with the correspondence indicated above between irreducible representation of G and those of G_1 and G_2. The Hilbert spaces on which the operators $R(g_1)$ and $S(g_2)$ act are denoted U, V, respectively, with dimensions d_u and d_v. We have $\Sigma\, d_u^2 = N_1$ and $\Sigma\, d_v^2 = N_2$ where the sums are extended over G_1 and G_2; also $d_u d_v = d_i$, the dimension of T_i. Now we work through the arithmetic in Equation (III.42), beginning with the inner sum. From the preceding paragraph we note $N_1 d_u^2 - d_u$ multiplications and $(N_1 - 1) \cdot d_u^2$ additions. Then to do the tensor product requires a further $N_2 d_u^2 d_v^2$ multiplications and $(N_2 - 1)d_u^2 d_v^2$ additions (working with matrix forms of the operators, where the tensor product goes over to a Kronecker product). Finally, we have to repeat this for all possible choices of R, S as we move through Γ_1 and Γ_2. Doing so over the 'R's first results in at most $N_2 N_1 d_v^2 + N_1^2$ multiplications; then counting all the S terms yields a final total of at most $N_1 N_1 N_2 + N_2 N_1^2 = N(N_1 + N_2)$ multiplications, to which we could add $r \leqslant N$ more because of the factor on the left side of Equation (III.42). Similarly we find a total of $N(N_1 + N_2 - 2)$ additions.

Thus the complexity of the group transform for general groups is about the same as for abelian groups of the same order. So, if the group factors into a product of more than two subgroups we can proceed just as before, down to the best result of $O(N \log_p N)$ operations, if the group permits that much factorization.

At this point the question of fast inverse transforms naturally arises. In the case of abelian groups it is clear, by duality, that the inverse transform is of the same complexity as the direct transform. That is, any factorization of G results in an analogous factorization of the dual group Γ, as indicated by Equation (III.41). The corresponding result for nonabelian groups is not so obvious, as the inverse transform (III.29) is not of the same form as the direct transform. Nevertheless, it has been shown [29] that the complexity, as we are measuring it, is the same for the inverse transform in the nonabelian case too.

From now on we have to be more specific to deal further with fast transforms. In considering a particular transform on some group G, we have to first see if G is decomposable in the sense that a factorization of the form (III.40) is possible and, if not, whether some other procedure can be effective. All of this circle of questions pertains to the subgroups structure of particular groups, and a successful resolution will require various additional assumptions about G.

Let's begin by stressing the role of cyclic groups. The cyclic group C_N of order N is abstractly defined by a single generator a and relation $a^N = e$. It is realized in various concrete ways as, for example,

(a) the subgroup of the circle group consisting of the Nth roots of unity;

(b) the quotient group Z/NZ, where Z is the integer group;

(c) the integers $0, 1, \ldots, N-1$ with addition mod N.

These are the most elementary examples of abelian groups, yet even these may well not be decomposable. Indeed, if p is a prime then C_p is a simple group, and C_pn is indecomposable (but not simple if $n \geqslant 2$). Hence, C_N is decomposable if $N = mn$ with $(m,n) = 1$.

We note an initial connection between representations and cyclic groups, realized as in part (a) above. Namely, let G be a group of order N and T a finite dimensional representation. Then for any $g \epsilon G$,

$$I = T(e) = T(g^N) = T(g)^N \quad ,$$

showing that the spectrum of T(g) is contained in C_N. In particular, if G is abelian then all characters on G assume values in C_N.

The cyclic groups are important in our subject primarily because those of prime power order are the building blocks of the general abelian group. In fact, the Fundamental Theorem for Abelian Groups permits us to describe all abelian groups of a given order. We recall the two-part statement, given an abelian group G of order N: first, if $N = p_1^{\alpha_1} \ldots p_n^{\alpha_n}$ is the prime factorization of N, then $G = G_1 \ x \ldots x \ G_n$ where G_i is the subgroup of elements of order p_i^t, $t \leqslant \alpha_i$; the order of G_i is $p_i^{\alpha_i}$, and this decomposition into subgroups of prime power order is unique. These subgroups G_i are called the Sylow subgroups of G. Second, each such 'p-group' is isomorphic to a product of cyclic p-groups. In fact, if H is any abelian group of order p^m, p a prime, then there is a unique list of integers $\{m_1, \ldots, m_r\}$, with $m_1 \geqslant \ldots \geqslant m_r \geqslant 1$, called the *type* of H, such that

$$H = C_pm_1 \ x \ldots x \ C_pm_r \quad .$$

This theorem is actually a special case of the cyclic decomposition of a finitely generated module over a principal ideal domain, e.g., [30, Chapter XV], but, of course, it can be established more directly, e.g., [30, Chapter I].

This theorem permits us to factor any finite abelian group into indecomposable (cyclic) factors and therefore to describe all abelian groups of a given order. For example, the following table (III-1) displays the distinct abelian groups of certain low orders in factored form, and also indicates the total number of (nonisomorphic) groups of that order.

TABLE III-1

Abelian Groups of Low Order

Order	Abelian Groups	Total Groups
4	$C_4, (C_2)^2$	2
6	C_6	2
8	$C_8, C_4 \times C_2, (C_2)^3$	5
9	$C_9, (C_3)^2$	2
10	C_{10}	2
12	$C_{12}, C_6 \times C_2$	5
16	$C_{16}, C_8 \times C_2, C_4 \times C_4, C_4 \times (C_2)^2, (C_2)^4$	14
24	$C_{24}, C_{12} \times C_2, C_6 \times (C_2)^2$	15
32	$C_{32}, C_{16} \times C_2, \ldots, (C_2)^5$	51
64	$C_{64}, C_{32} \times C_2, \ldots, (C_2)^5$	267(!)
96	$C_{32} \times C_3, \ldots, (C_2)^5 \times C_3$	230

With these preliminaries aside, let's now return to the subject of fast group transforms. We see that if the underlying group is abelian, the issue has been reduced to the case where the group is cyclic of prime power order. We know that such groups are indecomposable; nevertheless, they still have lots of subgroups. In general, if G is abelian, and m divides ord(G), then G has a subgroup of order m. This ails, however, for the simplest nonabelian groups. Now if G is cyclic and m divides ord(G) then there is exactly one (necessarily cyclic) subgroup H of G of order m and, in fact, this statement is characteristic of cyclic groups [31]. We can describe H explicitly in terms of any generator g of G: $H = \{e, h, \ldots, h^{m-1}\}$ where $h = g^k$, and $mk = \mathrm{ord}(G)$. Clearly, when G is cyclic of order p^n, the only possible values for $m = \mathrm{ord}(H)$ are $m = p^t$, $t \leqslant n$.

The following result now settles the specification of a fast algorithm for any group transform over a finite abelian group. Let G be abelian of order N and H a subgroup of order m and index s, so that ms = N. Then the complexity of the group transform on G is essentially $O[N(m + s)]$. To see this, we partition G into its cosets by H:

$$G = H \cup g_1H \cup \ldots \cup g_{s-1}H \quad ,$$

and then write, for each $\chi \in \Gamma$,

$$N\,\hat{f}(\chi) = \sum_{g \in G} f(g)\,\chi(g)$$

$$= \sum_{h \in H} f(h)\ \chi(g) + \sum_{h \in H} f(g_1 h)\ \chi(g_1 h) + \ldots$$

$$= \sum_{h \in H} [f(h) + f(g_1 h)\ \chi(g_1) + \ldots + f(g_{s-1} h)\ \chi(g_{s-1})]\chi(h) \quad . \tag{III.43}$$

From this expansion it is clear that the inner parenthesis involves (s - 1) multiplications and additions. Then we see (m - 1) more multiplications to accommodate the factors $\chi(h)$ (keeping in mind that $\chi(e) = 1$), and finally (m - 1) further additions. Repeating this for all N characters in G results in N(m + s - 2) multiplications and additions, as stated. Not surprisingly, this is the same figure as reported previously for decomposable groups. Naturally, this procedure can be repeated if H has a proper subgroup.

The preceding algorithm is basically the Cooley-Tukey method [27], designed originally for the ordinary DFT (about which more momentarily). Its extension to the abelian group context is due to Cairns [32], a few years later. During the following decade there was extensive development of fast algorithms for the DFT and other discrete transforms, based on number theory and matrix representations. So the method we have displayed, while providing a cute application of elementary group theory, and while of both historical and practical significance, is not the last theoretical word in computational efficiency. We have in mind especially the work of Winograd [33] and Nussbaumer/Quandalle [34], which is aimed at further reduction of the number of multiplications required to compute the DFT. Chapter 5 of [28] is a good general reference, while Reference [29] of Chapter II discusses the Winograd algorithm from an advanced standpoint. In the author's opinion, these fancier methods have not had a major impact in the day-to-day practice of computing large DFTs.

Looking in the other (chronological) direction we can note that the Cooley-Tukey procedure was not completely original or unprecedented. That their paper [27] had such an impact on digital signal processing was a matter of fortunate timing, reflecting both increasing appreciation of the uses of discrete Fourier and spectral analysis (physical chemistry, seismology, econometrics, etc.), and also increasing computing power. The historical record of the classical FFT includes the names of I. Good (1958), G. Danielson-C. Lanczos (1942), C. Runge (1903), and probably, it is fair to say, Buys-Ballot (1847). Some of this historical perspective is given in [35].

Having now referred to the classical DFT several times, beginning in Section II.1 and most recently just above, let us place this most prominent of discrete linear transforms into our group context. This is quite easy and, appropriately, it is associated with the most prominent of abelian groups: the cyclic group. Suppose we start with an arbitrary finite cyclic group C_N with generator a. Being abelian its irreducible unitary representations are all one-dimensional and coincide with its group characters. These comprise the dual group: $\hat{C}_N = \{\chi_1, \ldots, \chi_N\}$. Let

$$w_N = \exp(-2\pi i / N) \quad .$$

Then the formula

$$\chi_m(a^k) = w_N^{km} \quad , \qquad k = 0, 1, \ldots, N - 1$$

defines a character χ_m for each m = 1, . . . , N. It is easy to see that these characters are distinct from each other so that by the general theory we have all of them.

Now if $f = (f_0, \ldots, f_{N-1})$ is a function in $L^2(G)$, we apply the definition (III.24) of the group transform to obtain

$$\hat{f}(\chi_m) = \frac{1}{N} \sum_{k=0}^{N-1} f_k \, \chi_m(a^k)$$

$$= \frac{1}{N} \sum_{k=0}^{N-1} f_k \, w_N^{km} \quad , \qquad m = 1, \ldots, N \quad .$$

The right hand side here is recognizable as the usual formula for the DFT (e.g., [28, Chapter 3]). We have thus shown the ordinary DFT to be the group transform associated with a cyclic group of appropriate order. The standard properties of the DFT, its inverse, and the FFT now follow routinely from our more general group theoretic developments. Further, these developments also imply that the group transform on any finite abelian group is just a multidimensional DFT. Explicitly, given such a group G, we factor it as

$$G = C_{N_1} \times \ldots \times C_{N_s} \quad ,$$

where each N_i is some prime power and each C_{N_i} has generator a_i. The dual group Γ then factors similarly:

$$\Gamma = \hat{C}_{N_1} \times \ldots \times \hat{C}_{N_s} \quad ,$$

and each character on G is a product of characters on the associated cyclic groups. Hence the group transform on G has the form

$$\hat{f}(\chi) = \hat{f}[(\chi_1, \ldots, \chi_s)]$$

$$= \frac{1}{N} \sum_k f_k \, \chi[(a_1^{k_1}, \ldots, a_s^{k_s})]$$

$$= \frac{1}{N} \sum_k f_k \exp\left[2\pi i \left(\frac{k_1 m_1}{N_1} + \ldots + \frac{k_s m_s}{N_s}\right)\right] \cdot$$

where the summation index $k = (k_1, \ldots, k_s)$, each k_i runs from 0 to $N_i - 1$, and the indices $(m_1, \ldots, m_s)$ index the characters $\chi \epsilon \Gamma$, with $1 \leqslant m_i \leqslant N_i$.

In a slightly different direction, let's consider an integer N of the form $N = p^n$, p a prime. Let G be the corresponding group

$$G = (C_p)^n, \tag{III.44}$$

the product of n copies of C_p. We'll call such a group a *p-adic group*; in particular, a *dyadic group* when p = 2. Groups of this special structure are of frequent occurrence in applied mathematics. For example, they are exactly the possible additive groups of finite (Galois) fields, and hence are involved in treatmerts of algebraic coding theory. However, we shall refrain from such an excursion, and stick to signal theory here.

Suppose we realize C_p now as the set of integers $\{0, 1, \ldots, p - 1\}$. We can map G in a one-to-one manner onto the set $S_N = \{0, 1, \ldots, N - 1\}$ by the rule

$$x = \sum_{i=1}^{n} x_i \, p^{n-i}, \tag{III.45}$$

where each x_i This permits us to move back and forth between the group algebra $L^2(G)$ and vectors $f = (f_0, \ldots, f_{N-1}) \in C^N$ considered as functions on S_N. Now, as before, any character in Γ can be factored into a product of n characters, each in $\hat{C}_p$. Hence each character χ_m in Γ defines on S_N a function of the form

$$\begin{aligned} \phi_m(x) &= \prod_{i=1}^{n} \chi_{m_i}(x_i) \\ &= \prod_{i=1}^{n} w_p^{m_i x_i} \\ &= w_p \sum_{i=1}^{n} m_i x_i \quad , \end{aligned} \tag{III.46}$$

where the integers m_i and x_i are defined from m and x in S_N by the rule (III.45).

We have thus defined a family $\{\phi_m : m = 0, 1, \ldots, N - 1\}$ of functions on S_N, or, equivalently, a subset of C^N, which is seen to be a group-frame in C^N. Such a frame may be called a set of discrete Vilenkin-Chrestenson functions (cf. [36] and below). Two extreme cases are of special interest: the case $n = 1$, which is the DFT of prime order, and the case $p = 2$. This latter case involves the dyadic group of order 2^n. Here $w_p = w_2 = -1$, and so $\phi_m(x) = \pm 1$, for every m and x. (Of course, $|\phi_m(x)| = 1$ is true for all discrete V-C systems.) The frames B_1 and B_2 introduced at the beginning of Section III.2 are the special cases of this construction when $N = 4$.

The transform corresponding to the dyadic group case ($p = 2$) is usually referred to as the Walsh-Hadmard transform (WHT) [28, Chapter 8]. In matrix terms it is defined by

$$\hat{f} = H_{n-1} f \quad ,$$

where $H_{n-1} = \frac{1}{N} [u_{mk}]$, and

$$u_{mk} = (-1)^{\Sigma m_i k_i}$$

with the integers m_i, k_i, $i = 1, \ldots, 2^n$, once more coming from the rule (III.44). Thus, for e.ample,

$$H_1 = \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} \quad ,$$

and

$$H_n = H_{n-1} \otimes H_1.$$

This completes our examples of specific group transforms over abelian groups. Readers may consult the literature [1, 28, 36] for more details, many of which clearly follow, in unified fashion, from our group theoretical approach. Anď, as we recall from the end of Section III.1, there are many other unitary transforms which do not arise as group transforms.

Before leaving this topic we want to make a few final remarks. First, for any fixed N, consider the set of abelian groups of order N, and the associated group algebras which we have been using as sample spaces. Being N-dimensional these algebras are certainly equivalent as Hilbert spaces. But, as a consequence of a result of Kellogg [37] about certain commutative H*-algebras, they are actually isometrically *-isomorphic as H*-algebras. Therefore, the associated group transforms, which preserve all this structure, are strongly equivalent as representations of the algebras. Based on this observation, the structure theorem for finite abelian groups, and the fast algoritł. ıs which ensue from this structure, there would seem to be little basis for preferring one abelian group transform to another. Here is a specific instance of the general issue about data processing raised back in Section I.2: why do we perform one operation on data and not another?

The answer to this question must come from outside the mathematics; specifically, from the nature of the underlying signal and from the goal of the processing. For instance, the goal may be Wiener filtering or signal decorrelation. For filtering we refer to the next section. For decorrelation the signal statistics and performance criterion must be specified. An example of the latter might be

$$1 - \frac{r_s}{r_0} \quad ,$$

where r_s^2 is the quantity defined by Equation (III.10) and r_0^2 is the corresponding quantity for the original untransformed covariance matrix. Such measures of transform efficiency can be compared for varying group or other unitary transforms and data covariances. When the latter are taken to be those associated with Markov-1 processes, for instance, it is found that the DFT and WHT have similiar efficiencies for decorrelation (and also for signal coding), with a small advantage to the WHT [1, Chapter 3].

There is also the matter of the nature of the underlying signal from which, via sampling and other preprocessing steps, we have obtained our N-dimensional data vector. Although it would lead us too far afield to pursue this matter seriously, its importance requires us to at least indicate the issues. It is tied up with the problem of asymptotics already discussed in Section III.1. Here we will, as usual, emphasize the group-theoretic aspects.

We presume that our data sample has arisen from observations on some stochastic process indexed by an infinite group, such as the integers, the reals, etc. The covariance function of this process naturally determines the covariance matrix of the data vector, and there is nothing new to add to this aspect of the situation. Instead, what we want to think about is the nature of the sample functions. As already noted in Section II.3 there is little loss of generality in considering these functions to belong to a (weighted) L^2 space. There is then the question of an appropriate orthonormal coordinate system (or frame, again). We expect there to be a relation between the signal paths, a frame for the path space, and the finite frame which defines the unitary transform applied to the data vector. In particular, as a group transform is associated with a group-frame, we might ask about group-frames in the path space. This is really just the same issue as was raised early in this chapter except that here it is being extended to an infinite dimensional context. A complete theory will involve the suitability of the group-frame in signal space, which is basically a matter of approximation theory, and the expression of the attached group as a limit of finite groups — another kind of approximation. The suitability of a frame may also involve the statistical behavior of its coordinates as this derives from the assumptions about the underlying process.

As an example, suppose that this process is defined over a finite interval of the real line. Extending it periodically, we can view it and its sample functions as defined on the circle group T. Under its usual group structure T is compact abelian with dual group $\hat{T}$ isomorphic with the integer group Z by the correspondence $n \rightarrow \exp(int)$. Hence, expansion of functions in $L^2(T)$ is just classical Fourier series. Realizing the cyclic group C_N as a subset of T, the DFT appears as a sampled version of the usual Fourier expansion. This, of course, is well known. But T can be approximated by other finite groups. For instance, we can let $n \rightarrow \infty$ in the definition (III.44) of p-adic groups, and obtain

$$C_p^\infty = C_p \times C_p \times \dots ,$$

a compact abelian group in the product topology, with Haar measure equal to the product of the discrete Haar measures on each C_p. By using base-p expansions we can define almost everywhere on C_p^∞ a one-to-one transformation onto T which preserves Haar measure (recall that, on T, Haar measure is just normalized Lebesgue measure). If this transformation is employed to transfer the characters in (C_p^∞) over to $\hat{T}$ we obtain a frame in $L^2(T)$, the generalized Walsh functions of Chrestenson [38], with the ordinary Walsh functions resulting when $p = 2$.

More generally, we can take note of some work of Fine [39, I] who showed that in fact we can transfer the (countable) dual group of any compact metrizable abelian group into $L^2(T)$ so as to be an orthonormal set there, which is either finite or complete (and hence a frame). An analogous result remains valid for all spaces $L^2(P)$, P a probability measure on some measure space, if we assume that the orthonormal set, rather than the underlying space, carries a group structure [39,II].

The group C_p^∞ are examples of a special kind of topological group, called *Vilenkin groups*. These are, by definition, abelian topological groups that are second countable, periodic (each element belongs to a compact subgroup), and totally disconnected. Any compact Vilenkin group is essentially the direct sum $C_{p1} \times C_{p2} \times \ldots$, for certain primes $p_1, p_2, \ldots$. Such groups can be mapped onto T much as above, and their dual groups thereby go over into frames for $L^2(T)$. Because of the zero-dimensionality of a Vilenkin group, its dual group must be a torsion group; this special structure of the corresponding group-frames in $L^2(T)$ distinguishes them from general orthonormal systems there, and makes possible an elegant special theory of Vilenkin-Fourier series.

We conclude this section with a few remarks about the complexity of the group transform on a nondecomposable and nonabelian group. This is the only situation that has not been discussed so far, and naturally it leads to some deeper issues in group theory than have appeared to date. The passage from abelian to nonabelian groups should be viewed in the same spirit as, for example, the passage from stationarity to nonstationarity (cf. comments early in Section II.4).

Let G be a finite nonabelian group. If G is decomposable we can apply the analysis following Equation (III.42) to get a reduction in the complexity of the group transform F_G. If not, but there is a subgroup H, we can proceed in just the same way as was done in Equation (III.43) for abelian groups. If this is done it will be seen that the reduced operations count is essentially $O[N(m + s)]$ again (here $N = \mathrm{ord}(G)$, $m = \mathrm{ord}(H)$, $s = [G{:}H]$, as before). The problem now is that, unlike the abelian case where the indecomposable factors had the special form of cyclic groups of prime power order, the subgroup structure of general indecomposable groups is much more varied and complex.

To obtain the maximum reduction in complexity we would ideally like to find a subgroup H of maximal order in G, and then repeat this process in H, etc. This may indeed be possible for a particular G, but it is hard to generalize. There are two general ways to proceed: we can look at groups defined by special constructions (e.g., generators and relations, semidirect products), or by special properties (e.g., nilpotent, solvable), where at least the existence of an adequate number of subgroups may be guaranteed. In either approach our leitmotiv will be to only consider groups that are, in a suitable sense, 'close to abelian'.

At the outset it must be recognized that, again unlike the abelian case, if an integer m divides ord(G), there may not be a subgroup of order m. The standard example is the nonexistence of subgroups of order 6 in the alternating group A_4, which is of order 12. So it seems that we would like to restrict attention to groups G with the following property: if $\mathrm{ord}(G) = mn$ with $(m, n) = 1$ then G has a subgroup of order m. We encounter some good fortune here in that this property turns to be characteristic of a large and familiar class of groups, and further, this class is closed under the operations of forming subgroups, products, and quotient groups. The class in question is the class of all *solvable* groups. Such groups were historically first considered in connection with the problem of studying roots of a polynomial over a field; in this context and in a nutshell, a polynomial is solvable by radicals only if its Galois group is solvable, the general polynomial equation of degree n has the symmetric group S_n as Galois group and, for $n \geqslant 5$, S_n is not a solvable group.

There are various (equivalent) definitions of a solvable group. Perhaps the most familiar is that the (simple) factors of a composition series should be as 'simple' as possible, that is, abelian and hence cyclic of prime order This displays the sense in which solvable groups are close to abelian. How large is this class of finite groups? On the one hand, by the famous Feit-Thompson response (1963) to the classical Burnside conjecture, all groups of odd order are solvable. On the other hand, there is a great wealth of nonabelian simple groups (18 infinite classes, beginning with the alternating groups A_n, $n \geqslant 5$, and 26 additional 'sporadic' groups -- this is the Classification Theorem [40]), and so there is a correspondingly large number of nonsolvable groups than can, in principle, be constructed, say by Schreier's approach to the extension problem. Of course, most of these nonabelian simple groups have impractically large orders. Thus the smallest orders occurring are 60, 168, 360, 2520, 7920, ...

Specific examples of solvable groups of even order include those whose order = $2^m p^n$, where p is a prime and m, $n \geqslant 0$; in particular, groups of order 2^n.

A slightly more restrictive class of groups is especially convenient for optimally reducing the group transform complexity, for a given order. This is the class of nilpotent groups. Again, various definitions are possible. We give one motivated by subgroup structure. Suppose that G is a group of order N with prime factorization $N = p_1^{\alpha_1} \ldots p_n^{\alpha_n}$. By Sylow's theorems there is a subgroup of order $p_i^{\alpha_i}$ for each i = 1, . . . , n, and for each such i, all subgroups of order $p_i^{\alpha_i}$ are conjugate and hence isomorphic. These subgroups are the *Sylow subgroups* of G. Further, each Sylow subgroup contains a normal subgroup of all possible orders p_i^t, $1 \leqslant t < \alpha_i$. If, for each i, there is a unique Sylow subgroup, then G is the direct product of these subgroups, and conversely. Such groups are called *nilpotent.*

Additional details on the Sylow theorems, and nilpotent or solvable groups are available in standard sources; for example, the books by Hall [41], Hungerford [42], MacLane-Birkhoff [43], etc.

For applications, of course, we need some specific examples of nonabelian groups, preferably, as we have just indicated, those being nilpotent or at least solvable. We indicate next a few such examples. Among the most familiar examples of nonabelian groups are the symmetric group S_n and its (normal) subgroup A_n, the alternating group. We earlier noted that S_n is only solvable for $n \leqslant 4$. The groups S_n are interesting because of Cayley's classical result that any group of order N is isomorphic to a subgroup of S_N (via the right regular representation), while the A_n are interesting as the 'simplest' examples of nonabelian simple groups (for $n \neq 4$).

Keeping with our theme of only considering groups that are close-to-abelian, we consider next semidirect products of abelian groups. The simplest of these cases arises when the abelian groups are cyclic; the semidirect product of two cyclic groups is called a *metacyclic* group. The most familiar examples of this construction are the *dihedral* groups D_n (the symmetry groups of the regular n-gons). Given the cyclic groups C_m, C_n, the group generated by two elements a, b, satisfying the relations

$$a^n = b^m = e \quad , \quad b\, a\, b^{-1} = a^k \quad ,$$

with $k^m \equiv 1 \pmod{n}$, can be shown [43, p. 462] to be a semidirect product of C_m and C_n. The group D_n is the special case where $k = n - 1$, $m = 2$, and so $\mathrm{ord}(D_n) = 2n$. We find that D_2 is the (abelian) Klein 4-group, already mentioned in Section III.2, D_3 is the nonabelian group of least order, namely S_3, D_4 is the so-called octic group, etc. It turns out that dihedral groups are just those subgroups of an arbitrary group generated by a pair of distinct elements of order 2 (involutions), and that all subgroups of D_n are either cyclic or dihedral.

Since any extension of a solvable group by a solvable group is again solvable [43, p. 475], it follows in particular that metacyclic groups and, more generally, semidirect products of abelian groups are solvable.

The only other example of a nonabelian group of order < 10 is the *quaternion* group of order 8. This is the first nonabelian case of another class $\{Q_n\}$ of groups of order 4n called *dicyclic* groups. Q_n is defined abstractly as generated by a, b satisfying

$$a^{2n} = a^n b^2 = e \quad , \quad ab = ba^{-1} \quad .$$

The quaternion group Q_2 derives its name from its interpretation as $\{\pm i, \pm j, \pm k, \pm 1\}$ with group structure derived from the corresponding multiplication in the four-dimensional skew field of all quaternions. Q_2 may also be obtained isomorphically as the matrix group generated by the complex matrices

$$\begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix} \quad \text{and} \quad \begin{bmatrix} 0 & i \\ i & 0 \end{bmatrix} \quad .$$

(If we replaced i by 1 in the second matrix above, we would obtain D_4.)

Here we might make a quick comment intended to tie together some of this material with the general Heisenberg group concept mentioned at the end of Section II.4. Let p be a prime and F the field of p elements. The groups whose elements are of the form given in Equation (II.21), with a, b, $c \epsilon F$, is of order p^3. It turns out that there are only two distinct (nonisomorphic) groups of this order for any such p, and this construction easily defines one of them. In the case $p = 2$, this group is just D_4. Groups of this special nature are of interest in harmonic analysis because they furnish the simplest examples of asymmetry of the norms of convolution operators on the associated L^p spaces, $2 < p < \infty$ [50].

This completes our brief resume of relevant theory and examples of nonabelian groups. In practice, to apply the group transform methodology to signal coding or feature extraction, or to the group filters of the next section, we have to have available a complete list of irreducible unitary representations for the underlying group. That is, given the group G (abelian or not), we must know a representation T_i for each class comprising Γ, $1 \leqslant i \leqslant r$. Aid in this endeavor is provided by some facts presented in Section III.2, namely that r = number of conjugacy classes of G, and that the dimensions d_i of T_i divide ord(G) and obey the constraint $d_1^2 + \ldots + d_r^2 = \text{ord}(G)$. Then, in order to take advantage of possible savings in computational effort, we should also have available a composition series for G. Recall that this is a subnormal series (a descending sequence $\{H_i\}$ of subgroups with $G = H_0 \supset H_1 \supset H_2 \supset \ldots$, and each H_i normal in H_{i-1}) of maximal length, and that any two such series are equivalent by the Jordan-Holder theorem. Then the group transform can be efficiently computed, as we have shown, by nesting down the subgroups $\{H_i\}$.

III.5 GROUP FILTERS

This last section is somewhat more speculative than the others in this chapter and is, in fact, a major reason they were written. Our purpose is to define and briefly discuss group filters, and to examine one of their possible roles in data processing, namely, that of suboptimal Wiener filters. Unlike most of the foregoing material there is not a great deal of precedent for the use of general group filters in this fashion; we can cite only the recent note by Tracktenberg [44] which was, in turn, based on earlier joint work with Karpovsky [45]. That work was concerned with various deterministic problems centered around the use of group filters to approximate more generally defined systems, such as multivariate time-invariant linear systems defined over a finite discrete time interval. Once a group G has been chosen, this problem reduces to the approximation of an operator on $L^2(G)$, derived from the impulse response matrix of the given system, by group filters on G, in the Hilbert-Schmidt operator norm. Of course, the selection of G is both basic and difficult.

The definition of group filters is motivated by the description of transform coding in Section III.1, combined with the use cf group transforms. Namely, given a finite group G, and a data vector $f \epsilon L^2(G)$, we perform the following sequence of operations on f:

$$f \to \hat{f} = F_G(f) \to D\hat{f} \to F_G^{-1}(D\hat{f}) \quad , \tag{III.47}$$

where D is a diagonal operator on the transform domain. That is, D is defined on $L^2(\Gamma)$ by:

$$D(A, \ldots, A_r) = (A_1 \; D_1, \ldots, A_r D_r) \quad , \tag{III.48}$$

where each D_i is an operator on the d_i-dimensional space V_i of the ith irreducible representation T_i, in the notation of Equation (III.26). Thus the effect of the group filter is to selectively weight the spectral components $T_i(f)$, $i = 1, \ldots, r$, of the data. Extending the terminology of Pearl [7], a group filter is a linear basis-restricted transformation whose unitary component is a group transform.

Formally, then, a group filter is an operator on $L^2(G)$ of the form

$$F_G^{-1} \cdot D \cdot F_G \quad ,$$

where F_G is the group transform and D has the form displayed in Equation (III.48). An operations count reveals that such operators are generally of lower computational complexity than arbitrary operators on $L^2(G)$. Namely, if G is chosen so that F_G has a fast algorithm in the sense of Section III.4, and $N = \mathrm{ord}(G)$, then D can be computed with $d_1^2 + \, . . + d_r^2 = N$ multiplications, and, hopefully, the transform and its inverse can be done with $O(N \log N)$ multiplications each. Therefore, with reasonable choices of G, we can expect the multiplicative complexity of a group filter on G to be $N[1 + O(\log N)]$, compared with N^2 for a general operator.

From general properties of group transforms we see immediately that, alternatively, a group filter is simply a right convolution operator on $L^2(G)$. Such an operator sends f into $f * d$, where $T_i(d) = D_i$, $i \leqslant i \leqslant r$. (There is, of course, an equivalent theory of group filters involving left multiplication in Equation (III 48) and left convolution operators).

Having thus defined group filters in both the 'time domain' (convolution) and 'frequency domain' (multiplication), we can now state their *basic application*: for a given problem of discrete system simulation or of signal estimation, involving N-dimensional data, choose, for a specified group G of order N, an optima! group filter on G. The term 'optimal' is deliberately a little vague here, and an optimality criterion must be naturally be specified. Oftentimes, it is defined by a quadratic functional on $L[L^2(G)]$. Motivations for selecting group filters as approximating devices include speed of computation, suitability for specific architectures (especially filters on dyadic groups), and for specific signal statistics; see also [45].

Before getting into this topic in greater depth we want to offer just a few words of additional perspective here. Convolution operators on the standard infinite lca groups G, such as the circle group or R^k, are familiar and powerful tools of analysis. Much of their effectiveness is based on the concept of an approximate identity. This is a sequence $\}\delta_n\{$ in $L^1(G)$ with the property that for every $f \epsilon L^1(G)$,

$$f = \lim_{n \to \infty} f * \delta_n \; (= \lim_{n \to \infty} \delta_n * f)$$

in the metric of $L^1(G)$. Such sequences can be constructed on any of the standard lca groups by taking a bounded sequence $\}\delta_n\{$ in $L^1(G)$ with the properties that

$$\int \delta_n \, dmG = 1 \quad ,$$

and

$$\lim_{n \to \infty} \int_{G/V} \delta_n \, dm_G = 0 \quad ,$$

for each neighborhood V of e. For example, any sequence of probability density functions whose supports decrease to e would qualify, as would any sequence δ_n on R^k defined by

$$\delta_n(x) = n^k \phi(nx) \quad , \qquad \text{(III-49)}$$

for some fixed $\phi \epsilon L^1(R^k)$, $\int \phi(x)dx = 1$. Thus the Dirichlet and Fejer kernels are classical examples of approximate identities in Fourier series; the corresponding convolution operators are the partial Fourier series and Cesaro sum operators. In a different direction, the Gauss kernel $\phi(x) = (1/\sqrt{2\pi})\exp(-x^2/2)$ will serve in Equation (III.49); the corresponding convolution operators for positive real n solve the classical heat equation at time n. Another kind of convolution operator, derived from the Poisson kernel, solves the classical Dirichlet boundary value problem. So convolution operators are involved in many branches of analysis such as partial differential equations, potential theory, operator semigroups, Fourier analysis, etc. The point is that these familiar operators and approximation techniques are not part of our subject here. When the underlying group is finite, as is the case with discrete data processing, there is an identity for convolution; approximate identities are not required. Convergence concepts are not an issue; rather, it is the selection of high speed linear data processors for specific tasks, in a certain statistical environment.

Let's now look briefly at the nature of group filters, both individually and collectively, and then move on to their applications as suboptimal Wiener filters. We'll denote the set of all group filters on a given group G by $\Phi(G)$, and an element of $\Phi(G)$ with kernel ρ by T_ρ:

$$T_\rho(f) = f * \rho, \qquad f \in L^2(G) \quad .$$

Each operator T_ρ has its adjoint $(T_\rho)^* = T_{\rho^*}$. For both the operator norm $\|T_\rho\|$ and the Hilbert-Schmidt norm $\|T_\rho\|_2$, we have

$$\|\rho\| \leqslant \|T_\rho\| \leqslant \|T_\rho\|_2 \leqslant \|\rho\| \quad ,$$

and so both norms of the group filter T_ρ are equal to the norm of ρ as an element of $L^2(G)$. A group filter T_ρ is a normal operator whenever G is abelian or, more generally, whenever ρ is normal: $\rho * \rho^* = \rho^* * \rho$.

The set $\Phi(G)$ of all group filters on G is clearly a unital subalgebra of the entire operator algebra on $L^2(G)$. Indeed, it is the commutant of the set of left convolution operators, and conversely. (This relationship is, suitably interpreted, valid for general unimodular locally compact groups; see [46], and Reference [5] of Chapter II).

Next we show that the group filter algebra $\Phi(G)$ is N-dimensional, and exhibit a frame-basis for it. First, from the definition of convolution,

$$T_\rho f(g) = \sum_{h \in G} c_h \, R(h) \, f(g) \quad , \tag{III.50}$$

where $R(\cdot)$ is the right regular representation of G on $L^2(G)$, and $c_h = \rho(h^{-1})/N$. This, and the fact that each R(h) is of the form T_ρ, where $\rho = Ne_{h^{-1}}$, shows that $\Phi(G) = \text{span}\,\{R(h)\colon h \in G\}$. But, also

$$\text{tr}[R(h)] = \begin{cases} N, & h = e \\ O, & \text{otherwise} \end{cases}$$

and, therefore, since

$$\begin{aligned} \langle R(g), R(h) \rangle &= \text{tr}[R(g)\, R(h)^*] \\ &= \text{tr}[R(gh^{-1})] \quad , \end{aligned}$$

the operators $\{N^{-1/2} R(g)\colon g \in G\}$ constitute a frame in $\Phi(G)$.

(Another way to obtain the structure of $\Phi(G)$ is to simply note that the mapping $\rho \rightarrow T_\rho$ is a 1 anti-isomorphism from $L^2(G)$ onto $\Phi(G)$. It is norm-preserving and maps the frame elements $\sqrt{Ne_g}$, $g \in G$, of $L^2(G)$ onto the elements $N^{-1/2}R(g^{-1})$ of $\Phi(G)$, so these must constitute a frame there.)

It follows that the problem of system simulation, which reduces to the approximation of an operator on $L^2(G)$ by a group filter, is easily solved by an orthogonal projection of that operator on (G). In particular, it might be tempting to select a group filter for Wiener filtering by simply projecting the Wiener operator W of Equation (II.3) onto $\Phi(G)$. However, there is no reason to believe that this projection will minimize the mean square error over all group filters.

To obtain the best group filter for Wiener filtering we have two distinct, but ultimately equivalent, methods. In either case we are presented with a data vector $y \epsilon L^2(G)$, of the form

$$y = s + \eta$$

as per Equation (II.1), with known covariance operators C_s, C_η on $L^2(G)$. As usual, the signal s and the noise η are assumed uncorrelated. In the first method, we let T be an operator on $L^2(G)$. Then, for fixed s,

$$\begin{aligned} e(s;T) &= E(\|s - Ty\|^2) \\ &= \|s - Ts\|^2 + tr(TC_\eta T^*) \end{aligned}$$

Next, averaging over the signal prior distribution, we obtain the mean square error for each possible estimation operator T as

$$\begin{aligned} e(T) &= tr[(I - T)C_s (I - T)^* + TC_\eta T^*] \\ &= tr[T(C_s + C_\eta)T^*] - 2tr(TC_s) + tr(C_s) \\ &= T(C_s + C_\eta),T> - 2<T, C_s> + c, \end{aligned}$$

a quadratic functional of T.

Assuming the operator $C_s + C_\eta$ to be invertible (and, therefore, positive definite), the unconstrained minimum of $e(\cdot)$ over all T occurs at the Wiener filter $W = C_s(C_s + C_\eta)^{-1}$, as noted in Equation (II.3). Its mean square error was given in Equation (III.2). By contrast, the minimum of $e(\cdot)$ ovcr the subspace $\Phi(G)$ will be assumed at that group filter T_ρ for which

$$T_\rho(C_s + C_\eta) - C_s \perp \Phi(G) \quad . \tag{III.51}$$

If we express T_ρ as in Equation (III.50), the optimality condition (III.51) yields a system of linear equations for the coefficients $\{c_h;\ h \epsilon G\}$:

$$\sum_{h \epsilon G} a_{gh}\, c_h = b_g \quad , \quad g \epsilon G \quad , \tag{III.52}$$

where

$$b_g = <C_s, R(g)>$$

and

$$a_{gh} = <C_s + C_\eta, R(h^{-1}g)> \quad .$$

Thus, the matrix coefficients $\{a_{gh}\}$ of this system have the Toeplitz-like form $a_{gh} = \phi(h^{-1}g)$, where ϕ is easily seen to be a positive-definite function on G; the matrix $[a_{gh}]$ is therefore positive semidefinite. In fact, this matrix is positive definite, as long as the operatoı $C_s + C_\eta$ is invertible (which we are assuming).

Let T_w be the group filter whose coefficients in its expression (III.50) are the uniaue solution to the system (III.52). T_w is then our suboptimal Wiener filter. The impulse response function w is explicitly given by $w(g^{-1}) = Nc_g$, $g \epsilon G$. What is the additional mean square error incurred by the use of the filter T_w in place of the Wiener operator W? The answer is that

$$e(T_w) = e(W) + \langle W - T_w, C_s \rangle \quad ; \tag{III.53}$$

that is, the additional error is essentially the component of the difference operator along the signal covariance operator.

Various games can now be played with the right hand side of the resulting estimate

$$0 \leqslant e(T_w) - e(W) \leqslant \|W - T_w\|_2 \|C_s\|_2 \quad , \tag{III.54}$$

depending on what is assumed about the signal and noise statistics. For example, in terms of the normal power spectrum $\{\lambda_1, \ldots, \lambda_N\}$ of the signal s (equivalently, the spectrum of the covariance operator C_s), we have the bound

$$\begin{aligned} \|C_s\|_2 &= (\lambda_1^2 + \ldots + \lambda_N^2)^{1/2} \\ &\leqslant \lambda_1 + \ldots + \lambda_N = E(\|s\|^2) \quad , \end{aligned}$$

from Equation (III.9). If the signal is Gaussian then we can derive a sharp expression for $\|C_s\|_2$ by diagonalizing C_s and expanding s in the eigenvector frame $\{u_1, \ldots, u_N\}$:

$$\begin{aligned} s &= c_1 u_1 + \ldots + c_N u_N \quad , \\ \|s\|^2 &= |c_1|^2 + \ldots + |c_N|^2 \quad , \\ \|s\|^4 &= \sum_{i,j} |c_i|^2 |c_j|^2 \quad , \\ E(\|s\|^4) &= \sum_{i,j} E(|c_i|^2)\, E(|c_j|^2) \quad , \\ &= \sum_{i=1}^{N} E(|c_i|^4) + \sum_{i \neq j} E(|c_i|^2)\, E(|c_j|^2) \\ &= 3 \sum_{i=1}^{N} E(|c_i|^2)^2 + \sum_{i \neq j} E(|c_i|^2)\, E(|c_j|^2) \\ &= 2 \sum_{i=1}^{N} E(|c_i|^2)^2 + [\sum_{i=1}^{N} E(|c_i|^2)]^2 \quad , \end{aligned}$$

and, finally,

$$\|C_s\|_2^2 = \lambda_1^2 + \ldots + \lambda_N^2 = \frac{1}{2} \left(E\left(\|s\|^4\right) - E(\|s\|^2)^2\right) .$$

The other term on the right side of Equation (III.54) can be bounded by geometrical arguments: let $\{c_g\}$ be the coordinates of T_w in $\Phi(G)$, and let $d_g = <W, R(g)> / \sqrt{N}$, $g \epsilon G$, be the coordinates of the projection of W on $\Phi(G)$. Then one can check that

$$\|W - T_w\|_2^2 = \|W\|_2^2 + \|w\|_2^2 - 2 \sum_{g \epsilon G} \mathrm{re}(c_g \bar{d}_g) \quad ,$$

from two applications of the Pythagorean formula.

This completes our discussion of the first method for obtaining the optimum group filter T_w for Wiener filtering. We may refer to it as the 'direct method,' as it operates directly on the sample space $L^2(G)$ and associated operators. However, it is unsatisfying in that no formula for the filter T_w is obtained. We might indeed say that the solution is only indirectly presented via the Equations (III.52). A desire to remedy this difficulty, together with prior experience with deconvolution problems, leads us to the second 'indirect method.' In this approach we attempt to identify the optimum response function $w \epsilon L^2(G)$ rather than the operator T_w, and we do so by posing the problem as

$$w = \arg \min_{\rho \epsilon L^2(G)} E(\|y * \rho - s\|^2) \quad .$$

Letting now $e(\rho)$ denote the expectation on the right hand side above, averaged over first the noise and then the signal distributions, we have, after a Fourier transformation and some algebra,

$$\begin{aligned} e(\rho) &= E(\|y \cdot \hat{\rho} - \hat{s}\|^2) \\ &= \mathrm{tr}[\rho * (C_{\hat{s}} + C_{\hat{\eta}})\rho - 2C_{\hat{s}}\rho + C_{\hat{s}})] \quad , \end{aligned}$$

a quadratic function of $\rho \epsilon L^2(\Gamma)$. Its unconstrained minimum occurs at

$$\hat{w} = C_{\hat{s}}(C_{\hat{s}} + C_{\hat{\eta}})^{-1} \quad , \qquad \text{(III.55)}$$

a formula which is essenti:.lly given, without proof, in [44]. (In fact, we have chosen the notation C_s, C_η to agree, as much as possible, with that of [44]. We note that, for instance, C_s is that element of $L^2(\Gamma)$ whose ith value is the operator

$$N^{-2} \sum_{g,h} E[s(g)\overline{s(h)}]\, T_\iota(h^{-1}g) \quad ,$$

so that if P is the operator of right multiplication by A on $L^2(\Gamma)$, we have

$$E(\|P(\hat{s})\|^2) = <C_{\hat{s}} A, A> \quad .$$

Hence the optimal w in Equation (III.55) is computed as a product in the H*-algebra $L^2(\Gamma)$.

Thus, the indirect method gives the explicit formula (/II.55) for the transform of the optimal response function w, in terms of the signal and noise covariance components. Since the major point of using group filters as suboptimal Wiener filters is their reduced computational complexity, obtained by group transforming the data vector, multiplying by w, and inverse transforming, this method is to be preferred to the direct method because it makes w immediately available.

In analogy with Equation (III.2) one can verify that the minimal mean square error is given by

$$
\begin{aligned}
e(w) &= \sum_{i=1}^{r} d_i \, \mathrm{tr}[C_{\hat{s}} - C_{\hat{s}} \, (C_{\hat{s}} + C_{\hat{\eta}})^{-1} C_{\hat{s}})(i)] \\
&= \sum_{i=1}^{r} d_i \, \mathrm{tr} \left\{ [\hat{e}(i) - \hat{w}(i)] \cdot C_{\hat{s}}(i) \right\} \quad ,
\end{aligned}
$$

with $\hat{e}$ the identify in $L^2(\Gamma)$. Unfortunately, this does not seem as useful in assessing the increase in error over the Wiener filter as is the expression of Equation (III.53) and the subsequent estimate (III.54).

REFERENCES

1. R. Clarke, *Transform Coding of Images* (Academic Press, London, 1985).

2. T. Lynch, *Data Compression*, (Lifetime Learning Publications, Belmont, California, 1985).

3. S. Arnold, *The Theory of Linear Models and Multivariate Analysis* (Wiley, New York, 1981).

4. W. Pratt, "Generalized Wiener Filter Computation Techniques," IEEE **C-21**, 636-641 (1972).

5. I. Gelfand and A. Yaglom, "Calculation of the Amount of Information About a Random Function Contained in Another Such Function," Usph. Mat. Nauk **12**, 3-52 (1957) (Russian); AMS Translations **12**, 199-246 (1959).

6. M. Rosenblatt-Roth, "The Relative Entropy of a Random Vector with Respect to Another Random Vector," Tech. Report No. 85-85, University of Pittsburgh Center for Multivariate Analysis (1985).

7. J. Pearl, "Walsh Processing of Random Signals," IEEE **EC-13**, 137-141 (1971).

8. A. Segall, "Bit Allocation and Encoding for Vector Sources," IEEE **IT-22**, 162-169 (1976).

9. J. Pearl, "Asymptotic Equivalence of Spectral Representations," IEEE **ASSP-23**, 547-551 (1975).

10. Y. Yemini and J. Pearl, "Asymptotic Properties of Discrete Unitary Transforms, IEEE **PAMI-1**, 366-371 (1979).

11. M. Hamidi and J. Pearl, "On the Residual Correlation of Finite Dimensional Discrete Fourier Transforms of Stationary Signals, IEEE **IT-21**, 480-482 (1975).

12. ____________, "Comparison of the Cosine and Fourier Transforms of Markov-1 Signals," IEEE **ASSP-24**, 428-429 (1976).

13. N. Ahmed *et al.*, "Discrete Cosine Transform," IEEE **C-23**, 90-93 (1974).

14. M. Flickner and N. Ahmed, "A Derivation for the Discrete Cosine Transform," Proc. IEEE **70**, 1132-1134 (1982).

15. A. Jain, "A Sinusoidal Family of Unitary Transforms," IEEE **PAMI-1**, 356-365 (1979).

16. W. Ambrose, "Structure Theorems for a Special Class of Banach Algebras," Trans. Am. Math. Soc. 57, 364-386 (1945).

17. L. Loomis, *An Introduction to Abstract Harmonic Analysis* (D. Van Nostrand Co., Princeton, 1953).

18. R. Keown, *An Introduction to Group Representation Theory* (Academic Press, New York, 1977).

19. J.P. Serre, *Linear Representations of Finite Groups* (Springer-Verlag, New York, 1977).

20. R. Edwards, *Integration and Harmonic Analysis on Compact Groups* (Cambridge University Press, Cambridge, England, 1972).

21. M. Naimark and A. Stern, *Theory of Group Representations* (Springer-V :rlag, New York, 1982).

22. E. Hewett and K. Ross, *Abstract Harmonic Analysis II* (Springer-Verlag, New York, 1970).

23. L. Nachbin, "On the Finite Dimensionality of Every Irreducible Unitary Representation of a Compact Group," Proc. Am. Math. Soc. **12**, 11-12 (1961).

24. A. Lebow and M. Schreiber, "Polynomials over Groups and a Theorem of Fejer and Riesz, Acta. Sci. Math. **44**, 335-344 (1982).

25. J. Benedetto, "Harmonic Analysis and Spectral Estimation," J. Math. Anal. Appl. **91**, 444-509 (1983).

26. S. Kay and L. Marple, "Spectrum Analysis — A Modern Perspective," Proc. IEEE **69**, 1380-1419 (1981).

27. J. Cooley and J. Tukey, "An Algorithm for the Machine Calculation of Complex Fourier Series," Math. Computation **19**, 297-301 (1965).

28. D. Elliott and R. Rao, *Fast Transforms* (Academic Press, New York, 1982).

29. M. Karpovsky, "Fast Fourier Transforms on Finite Non-Abelian Groups," IEEE **C-26**, 1028-1030 (1977).

30. S. Lang, *Algebra* (Addison-Wesley, Reading, Massachusetts, 1965).

31. I. Richards, "A Remark on the Number of Cyclic Subgroups of a Finite Group," Am. Math. Monthly **91**, 571-572 (1984).

32. T. Cairns, "On the Fast Fourier Transform on Finite Abelian Groups," IEEE **C-20**, 569-571 (1971).

33. S. Winograd, "On Computing the Discrete Fourier Transform," Proc. Nat. Acad. Sci. USA **73**, 1005-1006 (1976).

34. H. Nussbaumer and P. Quandalle, "Fast Computation of Discrete Fourier Transforms Using Polynomial Transforms," IEEE **ASSP-27**, 169-181 (1979).

35. J. Cooley, P. Lewis, and P. Welsh, "Historical Notes on the Fast Fourier Transform," Proc. IEEE **55**, 1675-1677 (1967).

36. A. Trakhtman, "Fundamentals of the Linear Theory of Signals and Systems Defined on Finite Set of Points," Autom. Remote Control **4**, 589-599 (1974).

37. C. Kellogg, "Centralizers and H*-Algebras," Pac. J. Math. **17**, 121-129 (1966).

38. H. Chresterton, "A Class of Generalized Walsh Functions," Pac. J. Math. **5**, 17-31 (1955).

39. N. Fine, "On Groups of Orthonormal Functions I and II." Pac. J. Math. **5**, 51-65 (1955).

40. D. Gorenstein, "The Enormous Theorem," Sci. Am., 104-115 (Dec., 1985).

41. M. Hall, Jr., *The Theory of Groups* (Macmillan, New York, 1959).

42. T. Hungerford, *Algebra* (Holt, Rinehart and Winston, Inc., New York, 1974).

43. S. MacLane and G. Birkhoff, *A'gebra* (Macmillan, New York, 1967).

44. E. Trachtenberg, "Systems over Finite Groups as Suboptimal Wiener Filte. .. A Comparative Study", in *Mathematical Theory of Networks and Systems*, P. Fuhrmann, ed., (Springer-Verlag, 1984) pp. 856-863.

45. __________ and M. Karpovsky, "Some Optimization Problems for Convolution Systems over Finite Groups," Inf. Control **34**, 227-247 (1977).

46. I. Segal, "The Two-Sided Regular Representation of a Unimodular Locally Compact Group," Ann. Math. **51**, 293-298 (1950).

47. B. Fino and R. Algazi, "A Unified Treatment of Discrete Fast Unitary Transforms," SIAM J. Comput. **6**, 700-717 (1977).

48. H. Andrews, "Multidimensional Rotations ir. Feature Selection," IEEE **C-20**, 1045-1051 (1971).

49. J. Carl and C. Hall, "The Application of Filtered Transforms to the General Classification Problem," IEEE **C-21**, 785-790 (1972).

50. D. Oberlin, "$M_p(G) \neq M_q(G)$," Isr. J. Math. **22**, 175-179 (1975).

IV. CONCLUSIONS AND OUTLOOK

We will now summarize the foregoing material and emphasize some key points; along the way we will suggest a few promising directions of further research.

Chapter I represents an attempt to offer some general mathematical perspective on a very large and disparate field. It is deliberately presented at a low technical level, so as to be widely accessible. There is also an attempt to be just a bit provocative by designating a few results as being most fundamental from a mathematical viewpoint.

In a more serious vein we proposed a triangular array of mathematical areas as foundational for signal processing, namely, probability/statistics, Hilbert spaces/operator theory, and group representations/harmonic analysis. The value and interplay of the first two areas is by now familiar and well developed, and is not discussed herein in much detail. Let us just stress once again that such analysis begins with Equation (II-4), which we feel it fair to designate as the 'Fundamental Equation of Signal Processing'. According as the unknown signal x there is deemed to be deterministic or random, and based on the nature of the constraints and other prior information available concerning x, a variety of filters can be devised to optimize a particular performance measure. The Wiener and Gauss-Markov filters of Section III.1 are standard examples for recovering a random and a deterministic signal, respectively. In a different direction, the method of projection on convex sets (POCS) has become popular over the last few years. Here, all information about an unknown deterministic signal x (data + constraints) is combined to locate x in the intersection of a family of convex sets, and iterations involving the (generally nonlinear) projections on these sets are constructed to yield sequences which converge at least weakly, to the unknown signal [1]. The inherent nonuniqueness of these methods may be controlled by introducting a further cost functional [2].

The essential point here is that all these Hilbert space-centered methods have not been our major concern. We have rather chosen to study the role of our third foundational area: the group-related analysis. In doing so we eventually discerned three classes of application which could be indexed by the kind of group involved. Thus:

Group Type	Application
Finite	Digital signal processing (transform coding, pattern recognition, fast suboptimal filters)
Infinite lca	Weakly stationary and harmonizable signal models, filters, sampling
Heisenberg	Characterization of lca group transform; connnection with uncertainty principle and ambiguity function

The applications involving infinite groups were surveyed rather quickly in Sections II.2-II.5. Those involving the lca groups seem to have reached a mature stage, with work remaining at a fairly abstract level. It is worth emphasizing again that the group theoretic approach provides a systematic and unified approach to a frequency domain theory for weakly stationary processes, along with the associated invariant filters. By contrast, those applications involving the real Heisenberg groups and its representations are of quite recent development, are at a higher level of mathematical complexity, and are of somewhat more uncertain value. The connections with radar theory seem particularly worthy of further research efforts.

Finally, half of this report has been devoted to what appears to be the most promising area of immediate applications to the practice of digital data processing. The essential idea is the systematic use of those finite dimensional unitary transforms which can be realized as group transforms of some finite group. We noted in Section III.4 that, as a consequence of Kellogg's theorem, only nonabelian group transforms can be expected to significantly improve on the ordinary DFT in terms of error reduction in signal compression or filtering, although there may, for some purposes, be computational advantages which devolve from the group transform on some noncyclic abelian group (e.g., the Walsh-Hadamard transform on a dyadic group).

We can suggest some fairly natural research questions connected with this material and, indeed, Chapter III should be viewed as the necessary background and motivation for these questions, at the most elementary level. They all center around the association between a given covariance matrix (representing the second-order signal statistics) and the optimal group, of appropriate order, for a particular signal processing task. Having fixed such a task, such as signal decorrelation or Wiener filtering with additive white noise at a specified SNR, it is possible, in principle, to partition the cone of N × N positive semidefinite matrices into a finite number of subsets indexed by the appropriate optimal group. The number of subsets in this partition would equal 1 + the number of nonabelian groups of order N. The cone of matrices might be further reduced by imposing a bound on their norm (= spectral radius) or by requiring a constant diagonal (equal variances). For a fixed block length N, and signal processing task, this is itself a kind of pattern classification problem with the groups being the 'patterns'.

Suppose, in order to be specific, we consider the task of suboptimal Wiener filtering via group filters, as in Section III.5. Starting with the formula for Wiener filter W in Equation (II.3), and the error formulas (III.2), (III.53), we can readily derive the simple error relation

$$e\,(T_W) = \langle I - T_W, C_s \rangle,$$

for the optimal group filter T_W. This quantity can now be used as a performance measure to select the corresponding optimal group for each given signal covariance C_S. A few such studies for various data lengths N should reveal the potential of nonabelian groups and their associated transforms and filters to replace conventional methods based on the use of cyclic groups and the DFT.

The rather meagre evidence available to data, especially the computer experiment reported by Trachtenberg (Reference [44] of Chapter III) for the case where the signal s is derived from a first order Markov process, suggests that filters over nonabelian groups can indeed lower the mean square error of Wiener filtering by several percent over that determined by the DFT.

We might also emphasize here that group filters can serve to approximate any filter, not just the Wiener filter W. We have already noted that most filters derived by optimizing some Hilbert space performance criterion (Wiener, Gauss-Markov, maximum likelihood, projection filter, min-max, pseudoinverse, etc.) tend to be computationally intensive. However, such filters usually have finite dimensional domain, as the measurement operator A in Equation (II.4) is of finite rank when there is a finite number of observations, and so there is the general possibility of suboptimal approximations to each of these by a group filter. In the course of such investigations we might also expect to clarify the relative efficiencies of the direct and indirect methods of Section III.5 for actually obtaining group filters that optimize a particular performance measure.

In summary, we have presented an overview of existing and likely applications of group theory to various problems of signal processing and modeling. This effort is to be viewed as another of long ongoing series of group theory applications to various scientific and technical fields. The role of group theory in physics, chemistry, crystallography, etc., is, of course, one of long and honorable standing, tracing back to the seminal work of Weyl and Wigner. More recently we can see the infiltration of groups into statistical research [3, 4, 5]; this is in addition to the well-developed group role in time series models discussed in Chapter II above.

There is also a large body of material in the engineering literature that centers around the use of finite (Galois) fields. For present purposes we want to point out two just areas that are particularly related to the general theme of this report: number theoretic transforms and algebraic coding theory, especially group codes. The former are essentially group transforms defined on cyclic subgroups of the multiplicative group of a finite field, say GF(q). Naturally the lengths of such transforms are not arbitrary for a given integer q (necessarily a prime power), but are restricted to the divisors of q - 1. Such transforms can be used, together with their associated fast algorithms, to cyclically convolve integer sequences without round-off or overflow problems, and thus offer another approach to the fast FIR filtering and correlation of general real or complex data, after temporary rescaling. The recent survey by Blahut [6] provides a nice exposition of these ideas, and also discusses some issues of coding theory, such as the use of number theoretic transforms in a given field, finite or not, to define Reed-Solomon codes ('frequency domain coding').

The concept of group code was introduced by Slepian in 1956, and studied in a series of papers, of which we just cite [7, 8]. Originally, only binary channels were considered, and so a group code was defined as a subgroup of a dyadic group (in our terminology). With the geometric view that elements of the dyadic group of order 2^n correspond to vertices of the unit cube in real n-space, one could associate with a group code a (finite) subgroup of the n-dimensional orthogonal group which acts on the group code. The group code is an alphabet for describing the input to and output from the channel. The basic problem is to optimally encode data by means of the alphabet so as to minimize a mean square error criterion. This error will, of course, depend on the prior distribution over the data (this is often assumed to be uniform), and on the channel transition probabilities. Solutions to this problem (e.g., [9] and its references) utilize the dual group and group transform. This circle of problems has been extended to more general group codes [subspaces of arbitrary finite fields GF(q)] and, more recently, to the possible use of nonabelian group codes which, in some cases are already known to yield better performance than abelian group codes. We see here a strong parallel with the situations discussed in Section III.5 above, where nonabelian group filters show promise of outperforming the more conventional ones based on the DFT.

In conclusion, we have indicated many and varied applications of group theory and the associated harmonic analysis, at various levels of mathematical sophistication, to assorted engineering problems, primarily of a signal processing nature. We predict that group theory will eventually assume as fundamental a role here as, say, algebraic/differential geometry already has in the companion field of control theory.

Let us close by recalling the opinion of E. T. Bell, the well-known chronicler of mathematical history:

"Wherever groups disclosed themselves
Or could be introduced,
Simplicity and harmony
Crystallized out of comparative chaos. . . ."

REFERENCES

1. D. Youla and H. Webb, "Image Restoration by the Method of Convex Projections," IEEE, **MI-1**, 81-94 (1982).
2. R. Leahy and C. Goutis, "An Optimal Technique for Constraint-Based Image Restoration and Reconstruction, IEEE **ASSP-34**, 1629-1642 (1986).
3. M. Eaton, *Multivariate Statistics* (Wiley, New York, 1983).
4. R. Farrell, *Multivariate Calculations* (*Use of the Continuous Groups*) (Springer-Verlag, New York, 1985).
5. P. Diaconis, *Group Theory in Statistics* (Institute of Mathematical Statistics), to appear.
6. R. Blahut, "Algebraic Fields, Signal Processing, and Error Control," Proc. IEEE **73**, 874-893 (1985).
7. D. Slepian, "A Class of Binary Signaling Alphabets," Bell Syst. Tech. J. **35**, 203-234 (1956).
8. ______________, "Group Codes for the Gaussian Channel," Bell Syst. Tech. J. **47**, 575-602 (1968).
9. T. Crimmins, "On Encoding and Decoding Maps for Group Codes," IEEE **IT-22**, 763-764 (1976).

ACKNOWLEDGMENT

It is a pleasure to thank the entire secretarial staff: C.A. Poirier, D.M. McTague, and L.J. Parr, of Group 32, for their assistance in the preparation of this manuscript.

REPORT DOCUMENTATION PAGE

1a. REPORT SECURITY CLASSIFICATION
Unclassified

1b. RESTRICTIVE MARKINGS

2a. SECURITY CLASSIFICATION AUTHORITY

2b. DECLASSIFICATION/DOWNGRADING SCHEDULE

3. DISTRIBUTION/AVAILABILITY OF REPORT
Approved for public release; distribution unlimited.

4. PERFORMING ORGANIZATION REPORT NUMBER(S)
TR-761

5. MONITORING ORGANIZATION REPORT NUMBER(S)
ESD-TR-87-061

6a. NAME OF PERFORMING ORGANIZATION
Lincoln Laboratory, MIT

6b. OFFICE SYMBOL *(If applicable)*

7a NAME OF MONITORING ORGANIZATION
Electronic Systems Division

6c. ADDRESS *(City, State, and Zip Code)*
P.O. Box 73
Lexington, MA 02173-0073

7b. ADDRESS *(City, State, and Zip Code)*
Hanscom AFB, MA 01731

8a. NAME OF FUNDING/SPONSORING ORGANIZATION
U.S. Army Strategic Defense Command

8b. OFFICE SYMBOL *(If applicable)*

9. PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER
F19628-85-C-0002

8c. ADDRESS *(City, State, and Zip Code)*
U.S. Army Strategic Defense Command — Huntsville
Sensors Directorate
P.O. Box 1500
Huntsville, AL 35807-3801

10 SOURCE OF FUNDING NUMBERS

PROGRAM ELEMENT NO	PROJECT NO	TASK NO	WORK UNIT ACCESSION NO.
63220C, 63304A			

11. TITLE *(Include Security Classification)*
Mathematical Foundations of Signal Processing
II. The Role of Group Theory

12. PERSONAL AUTHOR(S)
Richard B. Holmes

13a. TYPE OF REPORT
Technical Report

13b. TIME COVERED
FROM ______ TO ______

14. DATE OF REPORT *(Year, Month, Day)*
13 October 1987

15. PAGE COUNT
108

16. SUPPLEMENTARY NOTATION
None

17. COSATI CODES

FIELD	GROUP	SUB-GROUP

18 SUBJECT TERMS *(Continue on reverse if necessary and identify by block number)*

signal processing	Wiener filter	group representation
finite groups	fast unitary transform	positive definite function
group transform	spectral coordinates	transform complexity
group filter		

19 ABSTRACT *(Continue on reverse if necessary and identify by block number)*

Several aspects of group theory that prove useful for various signal processing applications are presented.

Chapter I begins with a discussion of signal processing activities and goals at an abstract level, and continues with a look at the mathematical underpinnings of this subject. There follows a list of specific mathematical results that seem to be of greatest relevance to signal processing.

Chapter II surveys the role played by infinite groups in modeling signals and filters. Here substantial use is made of the associated harmonic analysis; in the abelian case the dual group serves as the natural frequency domain.

Chapter III presents a fairly detailed review of the representation theory of finite groups, through the Plancherel formula. The essential idea here is to then use those special unitary transforms which are also group transforms for digital signal compression and decorrelation, and the associated group filters as fast suboptimal Wiener (or other) filters. Initial evidence suggests that nonabelian group filters can impove on the standard DFT/FFT methods without significant increase in computational complexity.

Chapters I and III are written at an elementary level for wide access; Chapter II is written at a higher level, requiring some background in functional and harmonic analysis Comments are inserted throughout to suggest various generalizations of the material under discussion.

Chapter IV contains a summary of the main points and conclusions, and suggests some directions for further research, particularly on the use of finite nonabelian group transforms and filters

20. DISTRIBUTION/AVAILABILITY OF ABSTRACT
□ UNCLASSIFIED/UNLIMITED ☒ SAME AS RPT □ DTIC USERS

21 ABSTRACT SECURITY CLASSIFICATION
Unclassified

22a NAME OF RESPONSIBLE INDIVIDUAL
Capt. Arthur H. Wendel, USAF

22b TELEPHONE *(Include Area Code)*
(617) 863-5500, x-2330

22c OFFICE SYMBOL
ESD/TML

83 APR edition may be used until exhausted
All other editions are obsolete